

UNIVERSIDADE FEDERAL DE SANTA CATARINA

PROGRAMA DE PÓS-GRADUAÇÃO EM

ENGENHARIA DE PRODUÇÃO

***SIMULAÇÃO DE UMA CENTRAL DE ATENDIMENTO: UMA
APLICAÇÃO***

JUVENILLE FITZGERALD WOLFF

Orientador: Robert Wayne Samohyl, *Ph.D.*

Florianópolis, maio de 2003.

Juvenile Fitzgerald Wolff

***SIMULAÇÃO DE UMA CENTRAL DE ATENDIMENTO: UMA
APLICAÇÃO***

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do grau de Mestre em Engenharia de Produção.

Orientador: Prof. Robert Wayne Samohyl, *Ph.D.*

Florianópolis, maio de 2003.

*FICHA**CATALOGRÁFICA*

W855s Wolff, Juvenile Fitzgerald

Simulação de uma central de atendimento : uma aplicação / Juvenile
Fitzgerald Wolff; orientador Robert Wayne Samohyl. – Florianópolis:
2003.

68 f. : il. ; gráfs. ; tabs.

Dissertação (Mestrado) – Universidade Federal de Santa Catarina

Catálogo na fonte por: Onélia Silva Guimarães CRB-14/071

Juvenile Fitzgerald Wolff

***SIMULAÇÃO DE UMA CENTRAL DE ATENDIMENTO: UMA
APLICAÇÃO***

Esta dissertação foi julgada e aprovada para a obtenção do título de Mestre em Engenharia de Produção no Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina.

Florianópolis, 6 de maio de 2003.

Prof. Edson Pacheco Paladini, Dr.

Coordenador do Programa de Pós-Graduação em Engenharia de Produção

Banca Examinadora

Prof. Robert Wayne Samohyl, *Ph.D.*
Universidade Federal de Santa Catarina-UFSC
Orientador

Prof. Paulo José de Freitas Filho, *Dr.*
Universidade Federal de Santa Catarina-UFSC

Prof. Antônio Sérgio Coelho, *Dr.*
Universidade Federal de Santa Catarina-UFSC

Prof^a. Luciana Santos Costa Vieira, *MSc.*
Universidade do Sul de Santa Catarina-UNISUL

AGRADECIMENTOS

Ao meu orientador, professor Robert Wayne Samohyl pela sua contribuição e apoio na elaboração deste trabalho.

Ao Governo do Brasil e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-CAPES pela oportunidade e os recursos financeiros oferecidos durante o curso.

À Universidade Federal de Santa Catarina – UFSC pela disponibilidade de seus métodos, profissionais e recursos que contribuíram para o desenvolvimento deste trabalho.

Ao Prof. Avishai Mandelbaum do Instituto de Tecnologia do Israel pela sua autorização de usar os dados utilizados neste estudo.

Á todos que direta ou indiretamente contribuíram para a realização dessa pesquisa.

SUMÁRIO

1. INTRODUÇÃO	1
1.1. Objetivo Geral	2
1.2. Objetivos Específicos	2
1.3. Limitações e Delimitações	2
1.4. Estrutura do Trabalho	3
2. MODELAGEM E SIMULAÇÃO DE SISTEMAS	5
2.1. Definição da Simulação	5
2.2. Terminologia básica da simulação	6
2.3. Mecanismos de avanço de tempo	7
2.4. Incorporação da variabilidade dos sistemas reais aos modelos de simulação	8
2.5. Escolha de distribuições teóricas para representar os dados de entrada	9
2.6. Determinação do número de replicações	10
2.7. Passos de um projeto de simulação	13
2.7. Conclusão	16
3. CONCEITOS RELATIVOS ÀS CENTRAIS DE ATENDIMENTO	17
3.1. Definição de uma Central de Atendimento	17
3.2. Configuração de uma Central de Atendimento	18
3.3. Gerenciamento de Capacidade	19
3.4. Revisão da literatura voltada à simulação de Centrais de Atendimento	20
3.5. Métodos analíticos	21
3.6. Processo de chegada de pedidos de serviço	22
3.7. Medidas de desempenho para Centrais de Atendimento	24
3.7. Conclusão	25
4. MODELAGEM	26
4.1. A Central de Atendimento do Banco Anônimo	26
4.2. Descrição do Banco de Dados	27
4.3. Registros inconsistentes	29
4.4. Ajuste de Distribuições com o Arena Input Analyzer	30
4.5. Turnos e intervalos de trabalho	32
4.6. O Processo de Chegada de Chamadas	33
4.7. Tempo de serviço	36
4.8. Probabilidade de abandonos	37
4.9. Verificação e Validação do modelo	38
4.8. O Modelo de Simulação	40
4.8. Conclusão	42
5. EXPERIMENTAÇÃO	43
5.1. O modelo base	43
5.2. Comparação de Erlang-C com a Simulação	46
5.3. Análise em relação ao número de atendentes	48
5.4. Análise em relação ao volume de chamadas	50
5.5. Atendentes especializadas versus atendentes multifuncionais	51
5.5. Conclusão	53

6. CONCLUSÕES E SUGESTÕES	54
6.1. Conclusões	54
6.2. Recomendações e Sugestões	56
7. REFERÊNCIAS BIBLIOGRÁFICAS	57
ANEXOS	59

LISTA DE FIGURAS

Figura 1: Passos em um projeto de modelagem e simulação	13
Figura 2: Esquema operacional de uma Central de Atendimento	18
Figura 3: Comparação do número de chegadas do modelo com o sistema real.....	34
Figura 4: Chegadas de Serviço de tipo PS aos domingos	35
Figura 5: Chegadas do tipo NE aos domingos	35
Figura 6: Chegadas do tipo NW aos domingo	36
Figura 7: Fluxograma do processo de ligação da Central de Atendimento	41
Figura 8: Numero médio de chamadas por período	44
Figura 9: Comparação de Erlang-C com as observações do modelo e sistema real	48
Figura 10: Gráfico do tempo médio de fila para 4 diferentes intervalos de trabalho.....	50
Figura 11: O efeito do aumento do número de chegadas no tempo médio de fila	51
Figura 12: Tempo médio de fila do modelo base e o modelo alternativo	52

LISTA DE TABELAS

Tabela 1: Distribuições encontradas pelo <i>Arena Input Analyzer</i>	32
Tabela 2: Características de abandono de ligações no sistema real	38
Tabela 3: Teste-t para diferenças do tempo médio de fila	40
Tabela 4: Horário dos atendentes	44
Tabela 5: Tempo médio de fila através de Erlang-C e as observações no modelo.....	47

LISTA DE QUADROS

Quadro 1: Esquema de roteamento de chamadas	45
Quadro 2: Teste-t para a o tempo médio de fila das alternativas comparadas	49

LISTA DE ABREVIATURAS

BETA(Beta, Alpha): Distribuição de Probabilidades Beta com parâmetros Beta e Alpha
DAC: Distribuidor Automático de Chamadas
DISC: (CumP₁, Val₁,...,CumP_n, Val_n): Distribuição de Probabilidades Discretas
ERLA(ExpoMean, k): Distribuição Erlang com parâmetros ExpoMean e k
GNA: Gerador de Números Aleatórios
IID : Independentes e Identicamente Distribuída
Linf: Limite inferior do intervalo de confiança
LOGN(LogMean, LogStd): Distribuição Lognormal com parâmetros LogMean e LogStd
Lsup: Limite superior do intervalo de confiança
MCLM: Método Congruente Linear Multiplicativo
URA: Unidade de Resposta Audível
WEIB(Beta, Alpha): Distribuição Weibull com parâmetros Beta e Alpha

RESUMO

Simulação se encontra dentro da pesquisa operacional como a técnica com a maior aplicação no mundo empresarial. Com a evolução dos computadores e programas computacionais a simulação se apresenta como uma ferramenta ainda mais poderosa em comparação com alguns tempos atrás. Essa técnica se refere ao uso do computador para estudar uma variedade de modelos de sistemas reais através da avaliação numérica no computador utilizando um programa computacional para imitar o comportamento do sistema real.

Este trabalho apresenta uma análise de uma Central de Atendimento, através da simulação. Os dados utilizados são de uma pequena Central de Atendimento de um banco no Israel. Depois de estabelecer o objetivo deste estudo apresenta-se um resumo dos conceitos básicos da simulação. Em seguida é apresentada uma revisão da literatura voltada às Centrais de Atendimento. Na sequência, apresenta-se o modelo da central desenvolvido a partir do software Arena ® da Systems Modeling Corporation.

Na fase de experimentação são conduzidas análises tratando de intervalos de trabalho, estratégia de roteamento e aumento do número de chegadas de chamadas por unidade de tempo. Chega-se à conclusão que os intervalos de trabalho e aumento do número de chegadas de chamadas apresentam seu maior efeito no período de maior movimentação da Central de Atendimento enquanto, mudar a estratégia de roteamento de chamadas não provoca mudanças significativas no nível de serviço.

Palavras-chaves: Simulação, Modelagem, Centrais de Atendimento

ABSTRACT

Simulation is one of the most widely used techniques in operations and management science. Since computer and software are becoming better and cheaper simulation is more popular and powerful than ever. Simulation refers to the use of a computer to study a wide range of models of real world systems by numerical evaluation using software designed to mimic the behavior of the real system.

In this study a Call Center is investigated using simulation as an analysis tool. The source of the data used in this study is a small Call Center of one of Israel's banks. After the objective of the study is established an overview of some basic concepts of computer simulation is presented. In sequence a survey of some research related to the configuration of a Call Center is given. After the description of the data used in this research the Arena ® software of Systems Modeling Corporation is used to create a model of the entire call process. The model is then verified and validated to guarantee that it performs as intended. The model is then used to test sensitivities to the number of agents present to satisfy demand as well as changes in call volume. In conclusion the findings and the lessons learned from this study are presented in the final chapter.

Key words: Simulation, Modeling, Call Center

1. INTRODUÇÃO

A evolução da tecnologia de informação proporcionou maior disponibilidade de informações sobre processos na indústria e nos serviços, e a possibilidade de se analisar essas informações utilizando ferramentas quantitativas como a simulação, que até pouco tempo atrás era privilégio apenas das organizações de grande porte. Nos primeiros anos, quando surgiu a simulação, a técnica era usada nas grandes indústrias como último recurso para detectar a causa de problemas em linhas de produção. Nos anos 80, com a evolução dos computadores e a animação, a simulação começou a fazer parte das tomadas de decisão nas empresas. Em muitas empresas de grande porte, a simulação se tornou um requisito, antes da aprovação de grandes investimentos.

Nos anos 90, empresas de menor porte também adotaram a simulação no processo de tomada de decisão, e seu uso passou a ser utilizado nas primeiras etapas dos projetos. Com a maior disponibilidade de programas computacionais no mercado, ambientes cada vez mais amigáveis e melhor animação, a simulação se tornou uma ferramenta padrão em muitas empresas. A simulação também teve sua entrada na indústria de serviços onde é aplicada em muitas áreas não tradicionais, como por exemplo: hospitais, bancos e serviços de emergência (polícia, bombeiros, etc.).

A maior barreira para a simulação se tornar uma ferramenta universal e bem utilizada é o tempo necessário para o processo de modelagem e o treinamento especial para desenvolver um estudo de simulação.

Este trabalho demonstra a aplicação da simulação numa pequena Central de Atendimento de um banco em Israel. O banco de dados disponível na rede mundial de computadores (endereço: <http://ie.technion.ac.il/~serveng>) é de acesso livre. Centrais de Atendimento por telefone representam sistemas sujeitos a vários fatores aleatórios e fora do controle do analista. Consequentemente um modelo representativo de uma Central de Atendimento pode produzir resultados igualmente aleatórios necessitando-se assim de métodos estatísticos apropriados para fazer inferências sobre as verdadeiras características da Central. A simulação permite a inclusão destes fatores aleatórios e outros tipos de complexidade dos sistemas reais.

Em resumo pode se dizer que qualquer sistema industrial ou de serviços onde existe uma série de servidores e certo número de clientes aguardando numa fila para serem processados pode ser simulado.

1.1. Objetivo Geral

O objetivo geral deste trabalho é aplicar a metodologia da simulação numa Central de Atendimento de um banco comercial em Israel, através de métodos propostos na literatura e com auxílio do programa computacional do Arena.

1.2. Objetivos Específicos

A partir do objetivo geral proposto os seguintes objetivos específicos são estabelecidos:

- ✓ Obter uma visão melhor dos passos que envolvem o processo de desenvolvimento de um estudo de simulação e a análise dos resultados deste estudo.
- ✓ Levantar alguns parâmetros que possam influenciar o nível de serviço da Central de Atendimento.
- ✓ Levantar as dificuldades que se pode encontrar na aplicação de um estudo de simulação.

1.3. Limitações e Delimitações

Limitações:

- ✓ A falta de acesso direto aos gerentes da Central de Atendimento para verificar informações do banco de dados e a para obter dados mais detalhados.

- ✓ A própria versão acadêmica do Arena que apresenta um limite em termos de número de módulos que se pode incluir no modelo.

Delimitações:

- ✓ Apenas chamadas direcionadas aos atendentes são modelados considerando o horário de funcionamento da Central de Atendimento das 07:00 horas até as 24:00 horas.
- ✓ Não será abordada a otimização do sistema em termos de número de recursos mínimo para diminuir o tempo de espera na fila.

1.4. Estrutura do Trabalho

O trabalho está estruturado da seguinte maneira: o capítulo 1 traz a Introdução onde se apresenta uma idéia geral do assunto abordado neste trabalho.

No capítulo 2- Modelagem e Simulação de Sistemas- são apresentados conceitos básicos envolvidos na modelagem e simulação de sistemas. Apresentam-se, também, tópicos relativos à análise e ao tratamento dos dados de entrada de modelos de simulação e a análise dos resultados de um estudo de simulação.

No capítulo 3- Conceitos relativos às Centrais de Atendimento- é apresentado uma definição geral de uma Central de Atendimento, bem como alguns tópicos relativos a sua configuração e medidas de desempenho. É realizada, uma revisão da literatura voltada à simulação de Centrais de Atendimento. Apresentam-se também os métodos analíticos de cálculo de recursos e as suposições necessárias na sua aplicação.

No capítulo 4- Modelagem- é descrita a modelagem da Central de Atendimento: a escolha de distribuições e o processo de chegada e abandonos.

No capítulo 5- Experimentação- se apresentam experimentos com o modelo objetivando identificar variáveis que podem ter alguma influência significativa no modelo. As alternativas são comparadas com um modelo base baseado no domingo, o dia mais movimentado da semana.

No capítulo 6- Conclusões- são apresentadas as conclusões obtidas através da elaboração deste trabalho.

2. MODELAGEM E SIMULAÇÃO DE SISTEMAS

Este capítulo trata dos conceitos e principais técnicas envolvendo os programas e modelos de simulação. Inicia-se pela definição da simulação e em seguida trata-se de alguns conceitos básicos do método. Na seqüência, discute-se como a dinâmica e a variabilidade presentes nos sistemas reais são incorporadas aos modelos de simulação. Nos dois tópicos seguintes, apresentam-se conceitos associados à análise e tratamento dos dados de entrada de um modelo de simulação e os métodos estatísticos relacionados à análise dos resultados de um estudo de simulação. Finalmente, apresentam-se os passos envolvidos num projeto de simulação destacando a etapa da verificação e validação do modelo, por ser a garantia de que o programa computacional seja isento de erros e tenha um comportamento semelhante ao sistema modelado.

2.1. Definição da Simulação

A simulação é definida segundo Shannon (1998) como “um processo de projetar um modelo computacional de um sistema real e conduzir experimentos com este modelo com o propósito de entender seu comportamento e/ou avaliar estratégias para sua operação”. Um sistema segundo Shannon (1998) é “um conjunto de objetos, como pessoas ou máquinas, por exemplo, que interagem com a intenção de alcançar um objetivo ou um propósito lógico”. De acordo com Law e Kelton (1991), para estudar um sistema real geralmente precisa-se fazer várias suposições sobre seu funcionamento. De acordo com os autores, essas suposições normalmente transformadas em relações matemáticas ou lógicas, são constituídas num modelo com o objetivo de entender o comportamento do correspondente sistema. Este processo de modelagem exige uma série de simplificações sobre a organização e o funcionamento do sistema real com certo grau de abstração.

Se o sistema sob estudo apresenta alto grau de complexidade e uma solução analítica não é viável, a simulação muitas vezes é a única ferramenta para sua análise.

Na simulação com auxílio de um computador, o modelo é avaliado numericamente e dados são coletados para estimar as características verdadeiras do modelo (Law e Kelton, 1991). Se o modelo simulado possuir um ou mais elementos estocásticos, trata-se de uma simulação estocástica.

Como a razão principal de uso da simulação, a maioria dos artigos sobre o assunto cita, a possibilidade ao analista de investigar o que aconteceria se ele alterasse alguns parâmetros do sistema real sob estudo sem perturbar este sistema, uma vez que a análise é realizada no computador.

2.2. Terminologia básica da simulação

Na conceituação dos elementos básicos envolvidos na simulação utiliza-se a seguinte terminologia (Freitas, 2001):

Variáveis de estado. É aquele conjunto de variáveis necessárias para descrever o sistema num determinado instante no tempo, com relação aos objetivos do estudo.

Eventos. Um evento é uma ocorrência instantânea que pode causar uma mudança de estado no sistema. Em alguns modelos de simulação, eventos não necessariamente provocam mudanças no estado do sistema, mas apenas servem para programar outro acontecimento, como por exemplo, fim da simulação ou qualquer outra decisão sobre a operação do sistema (Law e Kelton, 1991). **Entidades e Atributos.** Um sistema é formado de um conjunto de entidades explicitamente definido. Neste sistema, uma entidade representa um objeto que pode ser dinâmica, movendo-se através do sistema, ou estática, servindo outras entidades. Atributos são características próprias das entidades. São dados que fazem parte das variáveis de estado do sistema.

Recursos e Filas de espera. Um recurso é uma entidade estática que fornece serviço às entidades dinâmicas. Uma fila de espera é uma coleção de entidades com alguma característica comum como, por exemplo, chamadas em espera numa central de atendimento. **Atividades e Tempo de Atraso.** Uma atividade é um período de tempo conhecido antes de sua execução. A duração pode ser uma constante, um valor

aleatório extraído de uma distribuição de probabilidades, o resultado de uma equação, de algum arquivo ou fonte externa ou calculada baseada no estado do sistema. Um tempo de atraso, ao contrário de uma atividade, é um período de tempo sobre o qual não se tem controle. Tempo simulado e replicações. O tempo simulado é o tempo real a ser simulado. Deve ser distinguido do tempo de simulação que representa o tempo necessário à execução de um experimento no computador. Uma replicação é a execução de um experimento com determinado tempo simulado. Cada replicação inicia e termina da mesma maneira e usa os mesmos valores para os parâmetros de entrada, mas utiliza diferentes números aleatórios para gerar os tempos entre chegadas das entidades e os tempos de serviço (Kelton et al., 1998).

2.3. Mecanismos de avanço de tempo

A simulação discreta consiste na modelagem de um sistema na sua evolução pelo tempo por uma representação onde as variáveis de estado do modelo mudam instantaneamente em pontos separados no tempo. Nas linguagens de simulação a natureza dinâmica do modelo é incluída por uma variável chamada de relógio da simulação que acompanha o tempo simulado enquanto a simulação avança (Freitas, 2001). Dois mecanismos para avançar o tempo podem ser utilizados: avanço do tempo para o próximo evento e avanço do tempo com incremento fixo. Neste trabalho, a referência será na primeira abordagem.

No mecanismo de avanço para próximo evento, o relógio da simulação é inicializado em zero e o tempo de ocorrência dos futuros eventos é determinado. Uma lista chamada de lista de eventos futuros é montada para o controle dos eventos já programados. O relógio é avançado para o tempo de ocorrência do evento mais eminente entre os futuros eventos da lista de eventos futuros. Neste momento, as variáveis do modelo e a lista de eventos futuros são atualizadas considerando o evento recém ocorrido. O relógio da simulação é novamente avançado para o tempo de ocorrência do próximo evento eminente da lista de eventos futuros e as variáveis de

estado e a lista de eventos futuros são atualizados. Este procedimento se repete até que alguma condição de parada predefinida seja satisfeita.

Segundo Freitas (2001) em relação à visão dos sistemas adotado na abordagem usando o mecanismo de avanço para próximo evento, as ocorrências de eventos futuros são explicitamente codificadas no modelo e são programadas para ocorrer no futuro simulado. De acordo com o autor uma abordagem diferente é a modelagem por processo onde os sistemas são modelados do ponto de vista das entidades e a codificação descreve a experiência das entidades se movendo através do sistema. Mesmo adotando a modelagem por processo a simulação é executada no computador na lógica do avanço para próximo evento (Law e Kelton, 1991).

2.4. Incorporação da variabilidade dos sistemas reais aos modelos de simulação

Nos modelos computacionais toda variabilidade necessária para a geração de dados é a função de um gerador de números aleatórios, o GNA (Kelton et al., 1998). O objetivo principal do GNA é gerar um fluxo de números aleatórios independentes e uniformemente distribuídos entre 0 e 1 (Freitas, 2001). O método adotado no Arena é o chamado Método Congruente Linear Multiplicativo (MCLM). Este método produz uma seqüência de números Z_1, Z_2, Z_3, \dots pela seguinte fórmula recursiva (Kelton et al., 1998):

$$Z_i = (aZ_{i-1} + c) \bmod m \quad 2-1$$

Onde m , a e c são constantes cujos valores devem ser escolhidos cuidadosamente baseados em critérios teóricos e empíricos para produzir uma boa seqüência de números aleatórios. A operadora $\bmod m$ significa dividir $aZ_{i-1} + c$ por m e passar o resíduo da divisão para o lado esquerdo da equação como o próximo Z_i (por exemplo: $422 \bmod 63$ é 44) (Kelton et al., 1998). Para inicializar o processo precisa se especificar

um valor Z_0 também chamado de semente. Finalmente para gerar valores entre 0 e 1 o último passo é dividir Z_i por m :

$$U_i = Z_i/m \quad 2-2$$

A seqüência U_1, U_2, U_3, \dots são os números pseudo-aleatórios necessários para o programa de simulação.

Como na fórmula recursiva **2-1** cada valor na seqüência depende apenas do seu antecessor é inevitável o modelo não apresentar ciclos de valores com periodicidade igual à constante m (Kelton et al., 1998). O valor de m depende da escolha dos parâmetros a e c . O GNA do Arena usa um MCLM com $m = 2^{31} - 1$, $a = 7^5 = 16.807$ e $c = 0$ e deve ser considerado suficiente para a maioria das aplicações (Kelton et al., 1998).

O GNA é a base da função nos programas de simulação que gera as variáveis aleatórias das distribuições teóricas usadas como dados de entrada do modelo (Kelton et al., 1998).

2.5. Escolha de distribuições teóricas para representar os dados de entrada

De acordo com Law e Kelton (1991) a maioria dos sistemas reais possui uma ou mais fontes de variáveis aleatórias e, para simular estes sistemas é preciso especificar a distribuição de probabilidades dessas variáveis. Conforme os autores, quando for possível coletar dados de uma variável aleatória de interesse estes dados podem ser usados em uma das seguintes abordagens para especificar a distribuição:

- 1) Usar os dados coletados do sistema diretamente na simulação sem qualquer tipo de tratamento.
- 2) Usar os dados coletados para definir uma distribuição de probabilidades empírica.

- 3) Usar métodos estatísticos para ajustar uma distribuição teórica aos dados coletados e através de testes de hipóteses analisar a aderência entre os dados reais e a distribuição escolhida.

Segundo os autores a primeira abordagem apresenta o defeito de produzir na simulação apenas valores históricos e, na prática, raramente se consegue dados suficientes para executar o número de replicações desejado. Kelton et al. (1998) ainda comentam que essa abordagem causa um atraso de processamento no computador devido à leitura dos dados durante a execução da simulação. A segunda abordagem, de acordo com os autores, evita estes defeitos desde que, pelo menos para dados contínuos, qualquer valor entre o valor mínimo e valor máximo das observações possa ser gerado na simulação. Eles recomendam a terceira abordagem se uma distribuição teórica pode ser encontrada com uma aderência adequada aos dados coletados. Os autores, entretanto, recomendam o uso do primeiro método para o processo de validação do modelo na comparação dos resultados do modelo de simulação com dados observados no sistema real.

2.6. Determinação do número de replicações

Como os valores sorteados de distribuições de probabilidades, objetivando a geração de dados voltados a alimentação de um modelo de simulação, são realizações particulares de variáveis aleatórias, estes valores podem apresentar grande variabilidade (Law e Kelton, 1991). Conseqüentemente, essas estimações numa única replicação podem variar significativamente em relação às verdadeiras características do modelo necessitando-se assim de métodos estatísticos apropriados para projetar e analisar experimentos com os modelos de simulação (Law e Kelton, 1991).

Dependendo da maneira de determinar o tempo simulado de uma replicação, classificam-se os modelos de simulação como terminais e não-terminais (Law e Kelton, 1991). Conforme Law e Kelton (1991) os modelos terminais são aqueles onde existe um evento natural que determina o tempo simulado de cada rodada de simulação

(replicação). Segundo os autores, as diferentes replicações usando números aleatórios independentes e sob as mesmas condições iniciais em cada replicação, implica que as variáveis aleatórias comparáveis entre si, nas várias replicações, possuem distribuições independentes e idênticas. Quando se deseja estimar o valor médio de uma variável de desempenho qualquer, no caso dos modelos terminais, a questão principal envolve a determinação do número de replicações mínimo necessário com um determinado nível de precisão e nível de confiança (Freitas, 2001). O cálculo do número de replicações mínimo é um processo iterativo iniciado pela execução piloto de n_0 replicações (Freitas, 2001). A relação entre o semi-intervalo de confiança e o número de replicações é dada pela equação (Law e Kelton, 1991):

$$h_0 = t_{n_0-1, 1-\alpha/2} \sqrt{\frac{S^2(n_0)}{n_0}} \quad 2-3$$

Onde,

n_0 : número de replicações piloto;

h_0 : semi-intervalo de confiança para n_0 replicações;

$S^2(n_0)$: variância para n_0 replicações;

$t_{n_0-1, 1-\alpha/2}$: valor crítico da distribuição t-student com nível de confiança $1-\alpha$, e n_0 graus de liberdade.

Resolvendo a equação 2-3 para n_0 :

$$n_0 = t_{n_0-1, 1-\alpha/2}^2 \frac{S^2(n_0)}{h_0} \quad 2-4$$

Se o valor crítico da distribuição *t-student* for aproximado pelo valor crítico $z_{1-\alpha/2}$ da distribuição normal padronizada e supondo que a estimação da variância $S^2(n_0)$ é aproximadamente igual para um maior número de replicações, n , pode-se estimar o número de replicações necessário para um semi-intervalo h e nível de confiança $1-\alpha$ ambos preestabelecidos pela fórmula, (Kelton *et al.*, 1998):

$$n \cong z_{1-\alpha/2}^2 \frac{S^2(n_0)}{h^2} \quad 2-5$$

Uma expressão mais simples é dada pela fórmula (Kelton *et al.*, 1998):

$$n \cong n_0 \frac{h_0^2}{h^2} \quad 2-6$$

O uso da distribuição *t-student* na fórmula 2-3 do cálculo do intervalo de confiança tem duas exigências (Kelton *et al.*, 1998): dados independentes e identicamente distribuídos e normalidade dos dados. A primeira exigência é garantida, como mencionado acima, pela utilização de diferentes números aleatórios para cada replicação na simulação. Segundo Kelton *et al.* (1998), o efeito no nível de confiança da violação da segunda exigência depende de vários aspectos incluindo a verdadeira distribuição dos dados e o número n , de replicações. O teorema central do limite estabelece que a soma ou a média resultante de um grande número de valores aleatórios e independentes é aproximadamente normal, independente da distribuição dos valores individuais (Kelton *et al.*, 2002). Segundo Kelton *et al.*, (1998) a pergunta natural é o que significa “um grande número de valores”. Os autores estabelecem que o efeito no nível de confiança depende da semelhança da distribuição dos dados da simulação à distribuição normal, particularmente em termos da simetria da distribuição. Quanto maior a assimetria na distribuição maior será o número de replicações necessário para obter um nível de confiança próximo àquela calculada pela fórmula 2-6.

Nos sistemas não terminais não se têm um estado inicial predefinido ao menos um evento caracterizando o fim do período de simulação (Freitas, 2001). Na análise dos resultados da simulação de sistemas não terminais as observações que são influenciadas pelo estado inicial do modelo devem ser desconsideradas desde que o interesse da análise esteja no desempenho do sistema quando o mesmo se encontra em regime (Freitas, 2001). O tempo de simulação deve ser de duração suficientemente longa para que os efeitos do estado inicial sejam esgotados (Kelton *et al.*, 1998). Se o tempo da fase transiente for pequeno em relação ao tempo da replicação pode se fazer a análise por meio de replicações independentes como no caso dos sistemas terminais descartando as observações da fase transiente (Kelton *et al.*, 1998). Outro método seria a realização de apenas uma simulação de duração suficientemente longa descartando as observações da fase transiente e dividir as demais observações numa série de lotes (Freitas, 2001).

2.7. Passos de um projeto de simulação

Um projeto de simulação consiste dos seguintes passos (Centeno *apud* Sangarayakul, 1998):

- 1) Formulação do problema
- 2) Formulação dos objetivos
- 3) Coleta de dados
- 4) Modelagem e codificação
- 5) Verificação e validação do modelo
- 6) Experimentação e análise
- 7) Documentação e recomendações

A relação entre os passos é demonstrada na Figura 1.

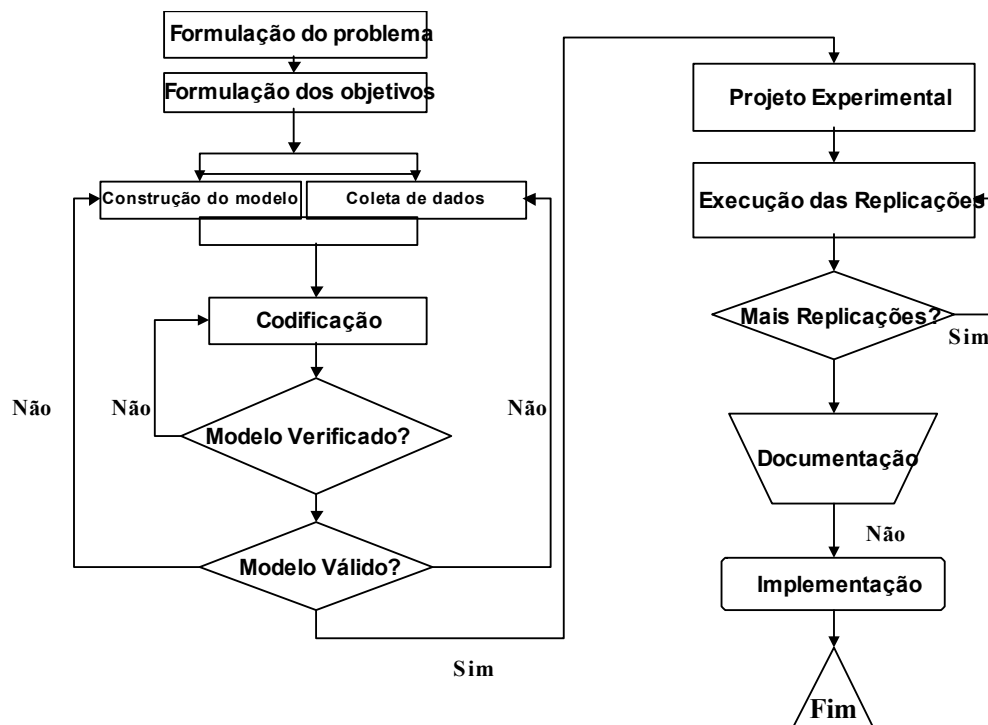


Figura 1: Passos em um projeto de modelagem e simulação

Fonte: SANGARAYAKUL, Bordin. *Use of Simulation to Analyze Block Manufacturing Methods*. (1998, pág.9)

Dos sete passos citados, a verificação e validação se destacam por garantir que os resultados obtidos pelo modelo de simulação e aqueles originados do sistema real apresentam uma proximidade razoável. Law *apud* Harris e Hoffman (1987) diz que “um dos mais importantes problemas enfrentando simuladores de sistemas reais é a tentativa de determinar se o modelo de simulação é uma representação suficientemente acurada do sistema sob estudo”. Verificação diz respeito ao processo de garantir que o modelo funciona conforme os pressupostos de modelagem estabelecidos (Kelton *et al.*, 1998). Para Kelton *et al.* (1998) validação consiste em garantir que o modelo se comporta de maneira semelhante ao sistema real. Kelton *et al.*, (1998) sugere as seguintes técnicas para o processo de verificação:

- “*Debugging*”: métodos de correção de programas computacionais
- Utilizar dados de entrada determinísticos que possibilitem determinar a resposta de saída do modelo com precisão
- Utilização de rotinas de rastreamento (*trace*) que permitem a visualização da lista dos eventos ocorridos

No caso da validação do modelo, Naylor e Finger (*apud* Law e Kelton, 1991) sugerem uma abordagem envolvendo três passos:

- 1) Desenvolver um modelo que pareça razoável aos indivíduos familiarizados com o sistema sob estudo. Para este fim, deve-se consultar todas as possíveis fontes de informação como: especialistas do sistema, dados levantados do sistema, teorias existentes e a própria experiência e intuição do analista.
- 2) Testar os pressupostos do modelo empiricamente. Neste passo deve se usar métodos quantitativos para testar as suposições feitas durante a fase inicial do projeto. Testes de aderência devem ser conduzidas para testar a

aderência das distribuições teóricas se forem utilizados, aos dados observados do sistema real.

3) Determinar o grau de representatividade dos resultados da simulação utilizando testes estatísticos. Law e Kelton (1991) recomendam dois métodos:

- a. O método de inspeção baseado na correlação positiva dos dados do modelo com os dados do sistema real. Eles recomendam comparar o sistema real e modelo executando o último com dados históricos coletados do primeiro, em lugar de distribuições teóricas e comparar os resultados dos dois.
- b. A construção de um intervalo de confiança baseado em dados independentes. Em casos onde for possível a coleta de um grande número de dados os autores recomendam a construção de um intervalo de confiança para diferenças através de um teste *t-student*.

Segundo Law e Kelton (1991), a comparação através de testes estatísticos sugerida pela literatura não é um processo simples, devido ao fato de que os resultados da maioria dos sistemas reais e seus modelos são não-estacionários e a existência de autocorrelação nas observações. Isto implica que testes estatísticos baseados em observações IID (Independentes e Identicamente Distribuída) não podem ser aplicados diretamente. Os autores ainda questionam se os testes de hipótese são apropriados desde que o modelo represente apenas uma aproximação do sistema real e a hipótese zero que o sistema e o modelo são "iguais" é certamente falso. Eles estabelecem que é mais apropriado responder se as diferenças entre o sistema real e o modelo são suficientemente significantes para questionar as conclusões baseadas do modelo.

2.7. Conclusão

Neste capítulo, além de algumas definições sobre a simulação e modelagem, foram apresentados os principais conceitos e técnicas envolvidos em programas de computadores de simulação. Foi apresentado como a dinâmica e aleatoriedade dos sistemas reais são incorporadas aos programas computacionais. Foi discutida também, a abordagem proposta na literatura para apresentar as variáveis de entrada do modelo. Em relação à análise dos resultados de um estudo de simulação foi discutido o método estatístico necessário para estimar qualquer valor de desempenho escolhido com uma precisão adequada. Finalmente foi apresentado um resumo dos passos envolvidos num projeto de simulação. A ênfase foi dada no processo de verificação e validação por ser a garantia de ter um modelo com comportamento semelhante ao sistema modelado. Isso permite a realização de inferências estatísticas a partir deste modelo com segurança. No capítulo a seguir, discute-se alguns conceitos relacionados às Centrais de Atendimento

3. CONCEITOS RELATIVOS ÀS CENTRAIS DE ATENDIMENTO

Este capítulo trata de conceitos relativos às Centrais de Atendimento. Inicia-se pela definição de uma Central de Atendimento. Em seguida, discute-se a configuração de uma Central de Atendimento simples e seu gerenciamento. Na sequência, apresenta-se uma revisão da literatura voltada à simulação de Centrais de Atendimento. Em sequência, são abordados os métodos analíticos de cálculo de recursos e os problemas encontrados na prática envolvendo as várias suposições necessárias na sua aplicação. Apresenta-se também o modelo do processo de chegada de chamadas proposto pela literatura. Finalmente, são apresentadas algumas medidas de desempenho para as centrais de atendimento.

3.1. Definição de uma Central de Atendimento

Segundo Koole e Mandelbaum (2001) uma Central de Atendimento é um conjunto de recursos, tipicamente pessoas, computadores e equipamentos de telecomunicação que possibilita a entrega de serviços por telefone. A Central de Atendimento é caracterizado pelo tipo de ligação: *inbound* ou *outbound*. Uma Central de Atendimento do tipo *Inbound* apenas recebe chamadas originadas externamente de clientes. No tipo *outbound* os atendentes efetuam as ligações aos clientes.

De acordo com Gans et al. (2002) a organização do trabalho nas Centrais de Atendimento pode variar de apenas um único nível - onde todos os funcionários estão envolvidos diretamente no atendimento de ligações externas - até múltiplos níveis - onde cada nível representa grupos de funcionários com habilidades especializadas - e os clientes podem ser transferidos pelos vários estágios do processo para completar o atendimento. Segundo os autores, quando a habilidade exigida for baixa, a empresa pode treinar cada funcionário a atender qualquer tipo de chamada e a ordem de atendimento pode ser de primeiro entrar primeiro a sair. Se existir a necessidade de

habilidades específicas, o funcionário especializado precisa de treinamento especial para atender um subconjunto de chamadas e as ligações são roteadas aos funcionários com habilidade apropriada para cada tipo de atendimento.

3.2. Configuração de uma Central de Atendimento

A Figura 2 demonstra um modelo operacional simplificado de uma Central de Atendimento. A configuração pode ser resumida da seguinte maneira (Gans et al., 2002): Um conjunto de k linhas telefônicas conecta as ligações à Central de Atendimento. Existem $w \leq k$ estações de trabalho onde $N \leq w$ funcionários atendem as ligações recebidas. Cada chegada de chamada encontrando todas as linhas ocupadas recebe o sinal de ocupado e é bloqueada antes de entrar no sistema.

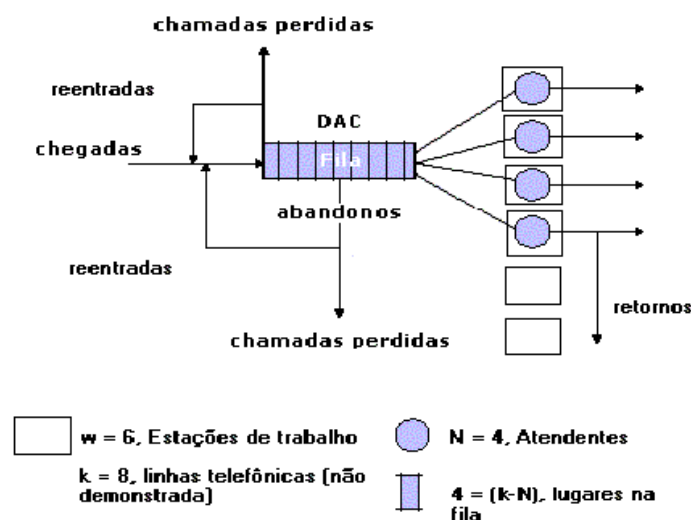


Figura 2: Esquema operacional de uma Central de Atendimento

Fonte: GANS et al.. *Telephone Call Centers: Tutorial, Review, and Research Prospects*. Technion, Out. 2002. Disponível em <http://ie.technion.ac.il/~serveng>. Acesso em: 19 dez. 2002.

No caso de encontrar uma linha livre, a ligação é conectada ao Distribuidor Automático de Chamadas, o DAC. Se existem menos que N atendentes ocupados no momento da entrada, a chamada é imediatamente encaminhada a um atendente. No caso de encontrar mais de N ligações no sistema, a chamada é retida na fila de espera. Clientes que ficam impacientes com a espera desligam ou abandonam antes de serem atendidos. Para os clientes que ficam esperando até o atendimento, a ordem de serviço é o primeiro a entrar e o primeiro a sair. A fração de clientes bloqueados que fazem novas tentativas, imediatamente após a primeira tentativa de ligação, representam as reentradas ao sistema. O restante de chamadas bloqueadas e abandonadas é perdido.

Finalmente, clientes atendidos também podem retornar ao sistema se não ficarem satisfeitos com o serviço ou para pedir serviços adicionais. O número de linhas telefônicas k atua como um limite superior para o número de chamadas contido no sistema, ambos em espera ou em serviço a cada instante. De forma semelhante o número de atendentes $N \leq w$, forma um limite superior para o número de chamadas servidas simultaneamente.

Ao decorrer do dia os gerentes da Central de Atendimento ajustam dinamicamente o número de atendentes para acompanhar a demanda de serviço. Com equipamentos adequados o número de linhas disponíveis k , também pode ser alterado. Por exemplo, para menor valor k pode se reduzir, durante o horário de maior movimento, o número de abandonos e tempo de espera (conseqüentemente os custos de ligação).

3.3. Gerenciamento de Capacidade

A questão da capacidade apropriada não deve ser limitada apenas ao número de atendentes e linhas telefônicas, mas envolve, também, todos os outros recursos como sistemas de informação e equipamentos de telecomunicação (Gans et al., 2002).

De acordo com Gans *et al.* (2002) operar uma Central de Atendimento próximo 100% de utilização, uma política geralmente praticada pelos gerentes de Centrais de Atendimento, dado o grande investimento em mão de obra, provoca maior tempo de espera na fila e o balanceamento destes dois fatores é a questão central na operação

de uma Central de Atendimento. Segundo eles, os custos de ligação para manter o cliente retido na fila de espera, especificamente nas Centrais de Atendimento de serviço gratuito, crescem aproximadamente linearmente com o número médio de clientes na fila. Nas Centrais de Atendimento que promovem vendas por telefone existe o custo de perda de oportunidade de vendas provocado pelo bloqueio de chamadas e abandonos (Andrews e Parsons, 1993).

3.4. Revisão da literatura voltada à simulação de Centrais de Atendimento

Chokshi (1999) relata como a simulação agregou valor como ferramenta de apoio à tomada de decisão no processo de reengenharia numa Central de Atendimento da AT&T. O autor comenta como a simulação ajudou a visualizar o processo e permitiu estimar as economias proporcionadas pelas mudanças.

Em um outro trabalho, Tanir e Booth (1999) relatam uma aplicação de simulação na Central de Atendimento da Bell Canadá. Neste trabalho o foco do modelo de simulação é a percepção do cliente na qualidade do serviço, modelando a experiência dele desde o início do serviço até o encerramento.

Klunge (1999) num artigo sobre a aplicação da simulação na Central de Atendimento de uma empresa de seguros comenta que a previsão é o passo mais importante no processo de gerenciamento de uma Central de Atendimento e os métodos tradicionais como Erlang-C, apesar de serem adequados para o planejamento geral, em casos onde suas suposições são violadas, apresentam a tendência de superestimar o número de atendentes. O autor recomenda o uso de simulação quando:

- Existem várias operações detalhadas;
- A precisão é importante;
- Existe grande variabilidade na demanda;
- Existe necessidade de animação para convencer a alta gerência;
- É necessário identificar gargalos no processo.

Harris *et al.* (1987) simulou a Central de Atendimento da IRS dos EUA e, através da simulação, chegou à conclusão que o processo, que funciona sob alto congestionamento, pode ser modelado de forma analítica. Segundo os autores, devido à natureza do serviço da IRS e limitações de recursos, o sistema sofre um grande número de reentrantes violando a suposição de chegadas de ligações conforme uma distribuição Poisson. O modelo resultante da pesquisa possibilita determinar, através de tabelas, a melhor configuração de atendentes e linhas telefônicas para atender a demanda prevista.

3.5. Métodos analíticos

Segundo Harris *et al.* (1987) na literatura de telecomunicações tradicionalmente o número de linhas telefônicas e o número de servidores necessários são calculados separadamente:

As fórmulas de Erlang-B são utilizadas para determinar o número de linhas telefônicas necessário para diminuir a probabilidade de bloqueio de ligações.

As fórmulas de Erlang-C são utilizadas para o cálculo do número de servidores necessário para diminuir a probabilidade de que uma ligação seja retida na fila de espera.

As suposições necessárias para o uso das fórmulas de Erlang-B são (Harris *et al.*, 1987):

- ✓ As chegadas ocorrem de acordo com uma distribuição de Poisson.
- ✓ Os clientes não fazem novas tentativas de reentrar no sistema logo depois de um bloqueio, mas aguardam, até que o sistema esteja menos congestionado.
- ✓ O número de linhas telefônicas pode ser calculado independentemente do número de servidores para atender as ligações.

As suposições necessárias para o uso das fórmulas Erlang-C são (Cooper, 1972):

- ✓ As chegadas ocorrem de acordo com uma distribuição de Poisson;
- ✓ Os tempos de atendimento possuem uma distribuição exponencial;
- ✓ O número de chegadas por unidade de tempo é menor que o número de pedidos atendidos por unidade de tempo;
- ✓ Existe um número ilimitado de linhas para reter as ligações não atendidas na hora de chegada;
- ✓ Não existem abandonos de ligações, ou seja, os clientes na fila de espera aguardam até o atendimento.

Gans *et al.* (2002) acrescentam ainda os pressupostos de um constante número de chegadas por unidade de tempo, ou seja, estacionaridade nas chegadas, constante número de ligações atendidas por unidade de tempo e a necessidade do sistema atingir rapidamente o estado de regime para cada intervalo sob análise.

Whittaker apud Harris *et al.* (1987) demonstrou que o cálculo do número de linhas telefônicas, separadamente do número de servidores, causa uma previsão equivocada de ambos o número de chamadas bloqueadas (por causa de falta de linhas) e o tempo médio dos clientes no sistema. Problemas com as suposições no uso das fórmulas de Erlang-C segundo Harris *et al.* (1987) aparecem em situações onde a capacidade de atendimento não é suficiente para atender a demanda e fazendo com que a terceira suposição não seja mais válida. No outro lado, um número ilimitado de linhas para reter ligações é fisicamente inviável.

3.6. Processo de chegada de pedidos de serviço

Segundo Harris *et al.* (1987), as primeiras tentativas de ligações telefônicas apresentam uma distribuição de Poisson desde que as chegadas num sistema telefônico não possam “enxergar” as demais chegadas e, por isso, não sejam afetadas pelo comportamento destas chegadas. Entretanto, os autores comentam que essa

suposição não é apropriada quando um grande número de clientes faz novas tentativas de entrar no sistema após um bloqueio (recebendo o sinal de ocupado) ou após abandonar a ligação (devido à impaciência de aguardar na fila de espera). Neste caso, o tempo de chegada de uma nova tentativa depende do tempo de chegada da tentativa anterior que seria uma violação da suposição da distribuição de Poisson.

Segundo Law e Kelton (1991) se $N(t)$ for o número de eventos a ocorrer antes do tempo t para $t \geq 0$, um processo estocástico $\{N(t), t \geq 0\}$ é um processo Poisson se:

1. As chegadas ocorrem apenas uma de cada vez;
2. $N(t + s) - N(t)$ (o número de chegadas no intervalo de tempo $[t, t + s]$) é independente de $\{N(u), 0 \leq u \leq t\}$;
3. A distribuição de $N(t + s) - N(t)$ é independente de t para todo $t, s \geq 0$.

Se $\{N(t), t \geq 0\}$ representa um processo Poisson, o número de chegadas em qualquer intervalo de s unidades de tempo é uma variável aleatória com parâmetro λs , isto é (Law e Kelton, 1991):

$$P[N(t + s) - N(t) = k] = \frac{e^{-\lambda s} (\lambda s)^k}{k!}$$

para $k = 0, 1, 2, \dots$ e $t, s \geq 0$

A média ou valor esperado é $E[N(s)] = \lambda s$ e particularmente $E[N(1)] = \lambda$, ou seja, λ é o número esperado de chegadas em qualquer intervalo de tempo com duração de uma unidade de tempo. Se $\lambda(t)$ for uma função do tempo t a terceira suposição não é mais válida e o processo é um processo de Poisson não estacionário.

Se $\Lambda(t) = E[N(t)]$ para todo $t \geq 0$ e $\Lambda(t)$ for uma função diferenciável para valores particulares de t , $\lambda(t)$ é definido como (Law e Kelton, 1991):

$$\lambda(t) = \frac{d}{dt} \Lambda(t)$$

Onde $\lambda(t)$ é a função de número de chegadas por unidade de tempo e $\Lambda(t)$ é a função de número esperado de chegadas no tempo t . De acordo com Law e Kelton (1991) o número de chegadas num intervalo $(t, t + s]$ para um processo de Poisson não

estacionário é uma variável aleatória com uma distribuição Poisson segundo o seguinte teorema:

Se $\{N(t), t \geq 0\}$ for um processo de Poisson não estacionário com uma função contínua de número esperado de chegadas $\Lambda(t)$, a probabilidade de k chegadas num intervalo de tempo $(t, t + s]$ é:

$$P[N(t + s) - N(t) = k] = \frac{e^{-b(t,s)} [b(t,s)]^k}{k!}$$

para $k = 0, 1, 2, \dots$ e $t, s \geq 0$

$$\text{onde } b(t, s) = \Lambda(t + s) - \Lambda(t) = \int_t^{t+s} \lambda(y) dy$$

Na prática $\lambda(t)$ [ou $\Lambda(t)$] deve ser estimado a partir de dados observados do processo de chegadas de interesse (Law e Kelton, 1991).

3.7. Medidas de desempenho para Centrais de Atendimento

A seguir apresentam-se algumas medidas de desempenho usadas nas Centrais de Atendimento. Entre essas medidas, o tempo médio de fila é escolhido neste trabalho como o tipo de desempenho que se deseja auferir no sistema.

Volume de Chamadas: o volume e a intensidade de pedidos de serviço são parâmetros importantes na determinação do número de recursos necessário numa Central de Atendimento (Diagnostic Strategies, 2000). É medida em erlangs: o produto do número de pedidos por unidade de tempo e o tempo médio de atendimento.

Período de Maior Movimentação (Diagnostic Strategies, 2000): é o período mais movimentado de uma hora de duração onde a probabilidade de que uma ligação seja bloqueada, abandonada ou retida na fila de espera é a maior. Tradicionalmente este é o volume de chamadas para calcular a quantidade de recursos necessários.

Tempo Médio de Atendimento (Diagnostic Strategies, 2000): define o tempo médio que um atendente necessita para atender uma chamada. Este tempo inclui o tempo de conversação e o tempo necessário para completar o serviço depois de encerrar a ligação.

Tempo médio na fila (Duder e Rosenwein, 2001): é o tempo médio que uma ligação deve esperar na fila para ser atendida.

Abandonos (Koole e Mandelbaum, 2001): é medida pela fração de clientes que abandonam a ligação antes do início do serviço.

Reentradas (Koole e Mandelbaum, 2001): é a fração de clientes nos quais os pedidos de serviço não são atendidos logo na primeira tentativa e voltam ao sistema para resolver o mesmo problema dentro de certo intervalo de tempo.

3.7. Conclusão

Neste capítulo apresentaram-se alguns conceitos de Centrais de Atendimento e tratou-se dos métodos tradicionais de cálculo do número de recursos necessário para atender a demanda de pedido de serviço. Apresentou-se também o modelo de chegadas de chamadas de acordo com um processo de Poisson estacionário e não-estacionário. No final do capítulo foram apresentadas algumas medidas de desempenho para centrais de atendimento. No próximo capítulo, apresenta-se o modelo de simulação da Central de Atendimento proposto neste estudo. Para isso, se utiliza a versão acadêmica do software Arena 3.0 que facilita a maior parte dos procedimentos dedicados à modelagem, verificação e validação de modelos e análise estatística de dados e resultados envolvendo um estudo de simulação.

4. MODELAGEM

Este capítulo é dedicado à modelagem da Central de Atendimento de um Banco comercial localizado em Israel e denominado ficticiamente neste trabalho, de Banco Anônimo. Nas primeiras partes do capítulo, apresenta-se o processo de ligação na central de atendimento do Banco e uma descrição dos dados usados neste estudo. Em seguida, apresenta-se o processo de modelagem que pode ser dividido em 4 partes principais: 1) modelagem em relação à configuração da central, incluindo o número de turnos e intervalo de trabalho, 2) modelagem do processo de chegada de chamadas, 3) modelagem do processo de serviço e 4) modelagem do processo de abandono de ligações. Em seguida, o tão importante processo de verificação e validação do modelo é apresentado o que deve garantir que o modelo de simulação tenha um comportamento semelhante ao do sistema modelado. Finaliza-se este capítulo com uma descrição sucinta do modelo de simulação.

4.1. A Central de Atendimento do Banco Anônimo

O Banco Anônimo presta serviço de atendimento aos seus clientes através de uma pequena Central de Atendimento. O serviço é diferenciado e trata de informações para clientes atuais e novos, informações de transações em conta corrente e poupança, informações de ações na bolsa de valores e assistência técnica para usuários da página do banco na internet. O horário de funcionamento é de 7:00 horas até meia noite nos dias de semana (Domingo até Quinta em Israel); para final de semana às Sextas, a central fecha as 14:00 horas e volta a funcionar aos Sábados as 19:00 horas até meia noite. Durante o horário de funcionamento, até 13 funcionários (13 estações de trabalho) podem estar presentes para fazer o atendimento incluindo um supervisor.

Pode se descrever o processo de ligação da seguinte maneira. Dependendo do serviço desejado o cliente liga para um dos números da central. Se não receber sinal de

ocupado o cliente é conectado à URA (Unidade de Resposta Audível) onde acontece sua identificação. Durante a interação com a URA o cliente recebe mensagens pré-gravadas e ele tem a possibilidade de auto-atendimento nesta etapa do serviço (a URA funciona 24 horas por dia para o auto-atendimento). Se ele em seguida não precisar de atendimento humano o cliente encerra a ligação e sai do sistema (6,19% dos clientes terminam o serviço nesta fase). Clientes que procuram falar com um atendente são encaminhados ao primeiro atendente disponível e o serviço é iniciado imediatamente. No outro caso a chamada é colocada em espera numa fila única.

Clientes na fila de espera são atendidos conforme a ordem de primeiro a entrar primeiro a sair. A posição na fila é determinada de acordo com o tempo de chegada na fila. A Central de Atendimento ainda utiliza um sistema de cliente preferencial subtraindo 1,5 minutos do tempo de chegada dos clientes preferenciais.

Enquanto retido na fila, o cliente recebe periodicamente informações sobre sua posição na fila. A gravação é repetida a cada 60 segundos com música, notícias e comerciais entre as repetições.

4.2. Descrição do Banco de Dados

O arquivo representa registros das chamadas atendidas na Central de Atendimento durante 12 meses do ano 1999. Para cada chamada, o registro tem 17 campos descritos a seguir: (uma amostra do banco de dados é demonstrada na tabela no Anexo).

- URA+Linha: Cada chamada é encaminhada à URA. Análise dos dados revela a presença de 2 URAs cada uma com capacidade de atender 16 chamadas ao mesmo tempo.
- No. de Identificação da Chamada: A cada chamada é atribuído um número identificador.

- No. de Identificação do Cliente: Este número representa a identificação do cliente e é zero nos casos onde o sistema não consegue identificar o cliente (como é no caso de clientes futuros).
- Prioridade: 0 e 1 indica clientes comuns ou não identificados e 2 indica clientes preferenciais
- Tipo: Indica o tipo de serviço pedido:
 - PS, PE - serviços gerais
 - IN - internet consulta
 - NE - informações sobre a Bolsa de Valores
 - NW - informações para futuros clientes
 - TT - para clientes que deixaram uma mensagem com o pedido ao banco para retornar sua ligação
- Data: Ano-mês-dia
- Horário de entrada na URA: A hora de chegada da chamada na Central de Atendimento.
- Horário de Saída da URA: A hora em que a chamada sai da URA: para entrar na fila, ser atendida diretamente ou devido um abandono.
- Tempo de Serviço URA: Tempo de atendimento na URA.
- Horário de entrada na Fila: A hora a chamada entra na fila. Igual a 0:00:00 para chamadas abandonadas ou servidas na URA.
- Horário de Saída da Fila: A hora em que a chamada sai da fila para o atendimento ou devido um abandono.

- Tempo de Fila: O tempo em que a chamada fica retida na fila de espera.
- Resultado: Existem três resultados:
 - *Agent* (atendente): chamada atendida
 - *Hang*: chamada abandonada
 - *Phantom*: uma chamada virtual ignorada neste estudo (os autores do banco de dados não encontraram uma explicação para este resultado e felizmente acontece com pouca frequência)
- Horário de início do Serviço: A hora de início do serviço prestado pelo atendente
- Horário de encerramento do Serviço: A hora de encerramento do serviço
- Tempo de Serviço: Tempo de serviço dos atendentes
- Servidor: Nome do agente que atendeu a chamada.

4.3. Registros inconsistentes

Existem registros de chamadas com tempo de serviço igual a zero enquanto o campo Servidor demonstra o resultado agente (12 ocorrências). Há registros com tempo de serviço maior que zero enquanto o resultado é “Hang”, ou seja, abandono (12 ocorrências). No mês de janeiro (o mês base de análise deste trabalho) estão faltando os registros de Segunda-feira 11 e Terça-feira 12.

Para janeiro, abril e julho aparecem chamadas de tipo AA (4 ocorrências), desconhecidas, já que o serviço AA não existe. Existem registros para qual o tempo de serviço ultrapassa o horário do início da próxima chamada, resultando em valores negativos para o tempo entre o fim e o início de chamadas consecutivas (poucas ocorrências).

Ao contrário da existência de 6 URA's, segundo a documentação relacionada ao banco de dados, os registros demonstram a presença de apenas duas URA's na Central de Atendimento.

4.4. Ajuste de Distribuições com o Arena Input Analyzer

A ferramenta computacional do Arena, o Arena Input Analyzer, possibilita a identificação de distribuições teóricas para representar no modelo os dados observados no sistema real através de testes de aderência e estimação dos parâmetros das distribuições encontradas. O programa fornece três medidas numéricas para quantificar a qualidade da aderência dos dados a uma distribuição teórica: o erro quadrado, o valor crítico e o valor p descritos a seguir.

Ordenando os dados em classes ou grupos de valores num histograma o Input Analyzer calcula o erro quadrado médio para o ajuste à distribuição teórica. O erro quadrado médio representa o valor médio das diferenças, tomado ao quadrado, entre os valores das frequências observadas nos dados amostrais e dos valores das frequências relativas da distribuição ajustada.

As duas outras medidas vêm de dois métodos estatísticos de aderência: o teste Qui-quadrado e o teste de Kolmogorov-Smirnov (K-S). Nestes testes as seguintes hipóteses são testadas:

H_0 : os dados seguem a distribuição teórica sob hipótese com os parâmetros estimados;

H_1 : os dados não seguem a distribuição teórica sob hipótese com os parâmetros estimados.

Para o teste Qui-quadrado, a estatística χ^2 é comparada com o valor crítico de $\chi^2_{\alpha, v}$ da distribuição de Qui-quadrado. A hipótese nula H_0 é rejeitada se $\chi^2 > \chi^2_{\alpha, v}$. Adicionalmente, o Input Analyzer fornece o chamado valor de p . Kelton *et al.* (1998) cita que “o valor p está associado à probabilidade de se obter um outro conjunto de dados

que seja mais inconsistente com a distribuição ajustada, do que o conjunto de dados atualmente utilizado. Maiores valores de p indicam maior aderência”. Segundo Freitas (2001) a literatura indica que para valores menores do que 0,05 a distribuição não é uma boa candidata. Por outro lado, se p for maior que 0,10 pode-se dizer que a distribuição teórica é uma boa representação dos dados reais.

Em casos onde não for encontrada uma distribuição teórica satisfatória o analista pode adotar uma distribuição empírica, baseada nos dados amostrais. O Input Analyzer define a distribuição empírica tomando a frequência relativa dos intervalos do histograma como a probabilidade de geração de valores daquele intervalo.

Segundo Kelton *et al.* (1998) não existe uma regra fixa de escolha de uma distribuição representativa. Os autores comentam que os testes estatísticos de aderência ou até a preparação dos dados (em termos de número de classes do histograma) podem ordenar as distribuições de maneira diferente. Law e Kelton (1991) ainda acrescentam que os testes não são suficientemente sensíveis para pequenas e até médias amostras e devem ser considerados apenas como um método sistemático para detectar grandes desvios. Segundo os autores, se a amostra for grande estes testes quase sempre rejeitam H_0 . Já que H_0 nunca é exatamente verdadeira até o menor desvio da distribuição hipotética que é detectado quando a amostra for grande, enquanto em geral para estudos de simulação é suficiente usar uma distribuição mais ou menos correta.

A Tabela 1 demonstra as variáveis aleatórias do modelo que precisam ser representadas por distribuições com suas respectivas representações teóricas encontradas com auxílio do *Arena Input Analyzer*. Observa-se que para amostras grandes a hipótese nula é rejeitada como foi visto acima. Para os objetivos deste estudo, entretanto, as distribuições encontradas podem ser consideradas boas representantes das correspondentes variáveis do sistema real e são, portanto, usadas no modelo.

Variável aleatória a ser modelada	Expressão no Arena	Erro Quadrado	Valor Critico	Valor P	Graus de Liberdade	Número de Observações
Tempo de Serviço na URA	10 + WEIB(3.92, 1.13)	0.000639	372	< 0.005	4	7227
Tempo de Tolerância	0.999 + EXPO(60.1)	0.002351	79.8	< 0.005	8	3080
Tempo de Pós-Ligação	Empírica					8080
Tempo de Serviço Chamada PS	10 + LOGN(193, 242)	0.000090	37.7	0.0097	20	19540
Tempo de Serviço Chamada PE	10 + GAMM(206, 1.31)	0.005667	3.22	0.213	2	111
Tempo de Serviço Chamada NW	10 + LOGN(357, 544)	0.002824	20.3	0.089	13	1397
Tempo de Serviço Chamada TT	10 + LOGN(108, 175)	0.000152	15.5	0.225	12	3729
Tempo de Serviço Chamada NE	10 + WEIB(124, 0.759)	0.000371	1.44	0.491	2	353
Tempo de Serviço Chamada IN	10 + LOGN(282, 397)	0.001342	2.52	0.478	3	197

Tabela 1: Distribuições encontradas pelo *Arena Input Analyzer*

4.5. Turnos e intervalos de trabalho

O número de agentes atendendo chamadas ao longo do dia é uma informação importante, entretanto, não diretamente disponível no banco de dados. Uma maneira de estimar o horário e número de atendentes por turno pode ser feito com auxílio de uma planilha em Excel® ou Access® filtrando para cada dia de semana o tempo do início de atendimento da primeira chamada e o tempo do fim de atendimento da última chamada atendida pelo atendente. Arredondar estes horários ao horário inteiro mais próximo dá uma aproximação dos turnos e o número de atendentes presentes naquele dia. Igualmente a duração dos turnos só pode ser uma aproximação do sistema real.

Outra informação importante para o modelo é o horário e a duração dos intervalos de descanso para os agentes. Essa informação também não é diretamente disponível no banco de dados e deve ser estimada. Uma maneira de estimar este dado pode ser o cálculo dos tempos entre o fim de atendimento de uma chamada e o início da próxima chamada atendida pelos agentes individuais. Este cálculo inclui os eventuais intervalos de trabalho, períodos sem chegadas de chamadas, e possíveis durações de tempo de trabalho após o encerramento da ligação para completar o serviço. Com os dados disponíveis no banco de dados, não é possível filtrar estes tempos para obter um maior granulação da informação. O tempo de pós-ligação deve ser estimado através deste cálculo do tempo entre chamadas consecutivas. Se estes valores são truncados,

eliminando valores maiores que 5 minutos para os atendentes com maior número de ligações atendidas pode-se obter desta maneira uma estimativa do tempo de pós-ligação. A Tabela 1 indica que uma distribuição empírica é escolhida para representar o tempo de pós-ligação no modelo.

4.6. O Processo de Chegada de Chamadas

Um pressuposto básico do modelo é que a chegada de chamadas pode ser representada por um processo de Poisson não homogêneo. Outro pressuposto é a existência de sazonalidade diária nos dados. A função de número de chegadas por unidade de tempo $\lambda(t)$ pode ser aproximada, dividindo o dia em 34 intervalos de 30 minutos e coletando vários valores para cada subintervalo, tomando a média como estimativa do número de chegadas daquele subintervalo. Para gerar as chegadas no modelo o método chamado “rejeição” (Lewis e Shedler *apud* Kelton *et al.*, 1998) é aplicado. Primeiramente o valor máximo do número de chegadas por unidade de tempo λ_{max} , é determinado. Em seguida um processo de Poisson estacionário utilizando λ_{max} é gerado para cada intervalo de 30 minutos. Cada chegada gerada, entretanto, é rejeitada com determinada probabilidade. A probabilidade de uma chegada gerada pelo o processo de Poisson estacionário entrar no sistema é igual ao número de chegadas de determinado período λ_i , dividido pelo valor máximo de chegadas λ_{max} , dos 34 períodos.

No sistema real, apenas 6,2% das chamadas são terminadas na URA (auto-atendimento) e não apresentam influência significativa no tempo de fila e são desconsideradas no modelo.

A Figura 3 demonstra o gráfico da função de chegadas por intervalo de meia hora de domingo até sábado gerada pelo modelo, juntamente com as médias das observações reais por intervalo de meia hora.

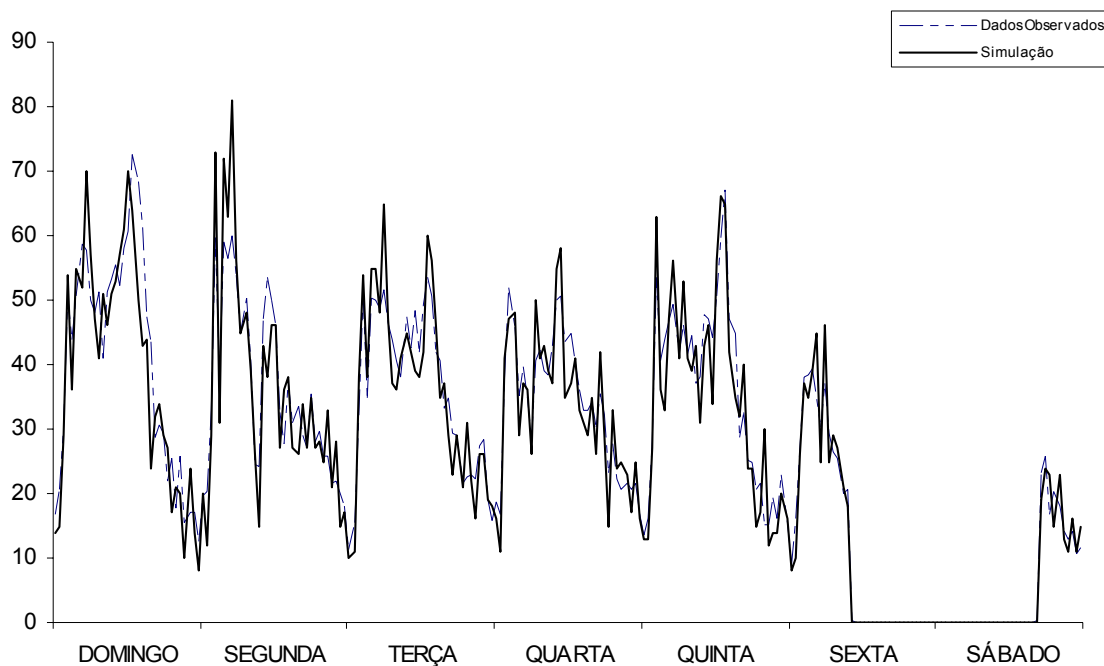


Figura 3: Comparação do número de chegadas do modelo com o sistema real

A proporção de chamadas com prioridade 2 (clientes preferenciais) é igual a 0,28 para o mês de janeiro. Uma análise mais precisa das chegadas, demonstra que o volume de chamadas por tipo de serviço também é não-estacionário. As Figuras 4, 5 e 6 demonstram o número de chamadas por período de duração de uma hora para respectivamente os serviços PS, NE e NW nos domingos do mês de janeiro. Pode-se observar que a sazonalidade diária nas chegadas é evidente. As demais chamadas de tipo TT, PE e IN não apresentam um padrão significativo ao longo do dia e representam as menores proporções de chegadas no sistema.

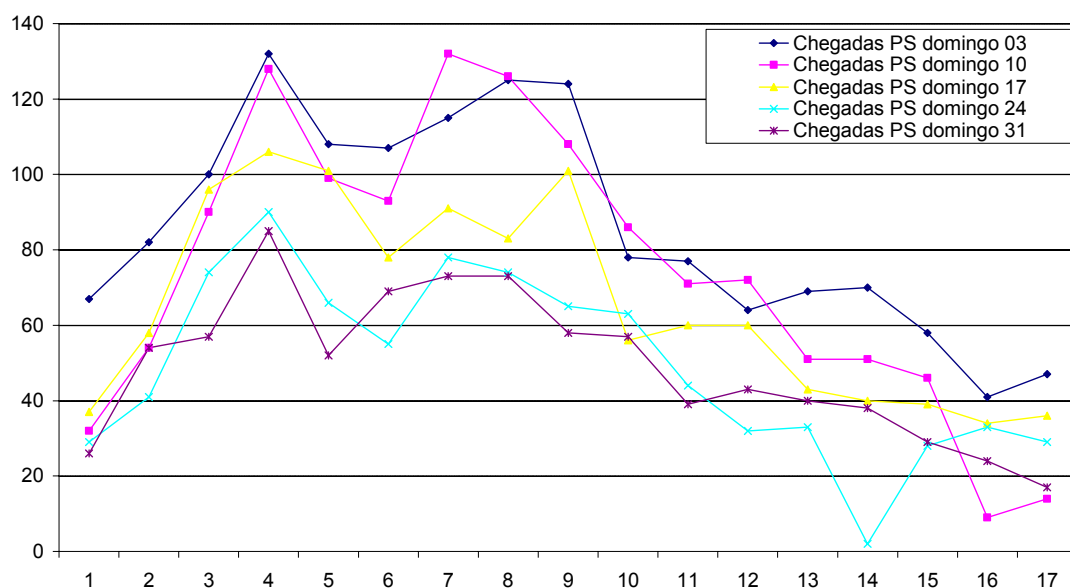


Figura 4: Chegadas de Serviço de tipo PS aos domingos

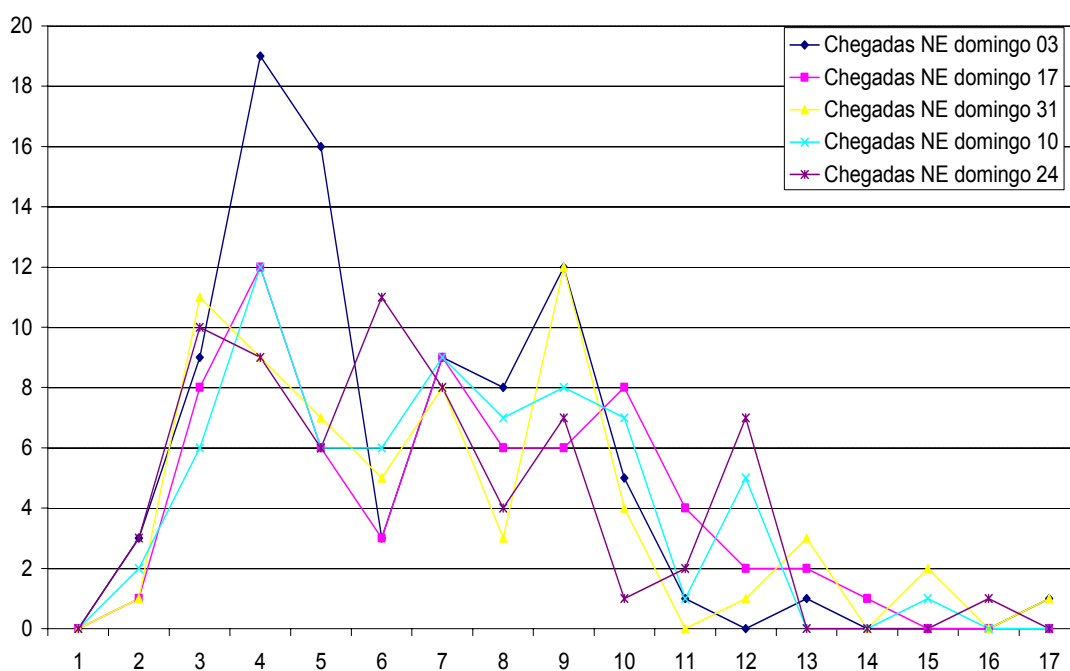


Figura 5: Chegadas do tipo NE aos domingos

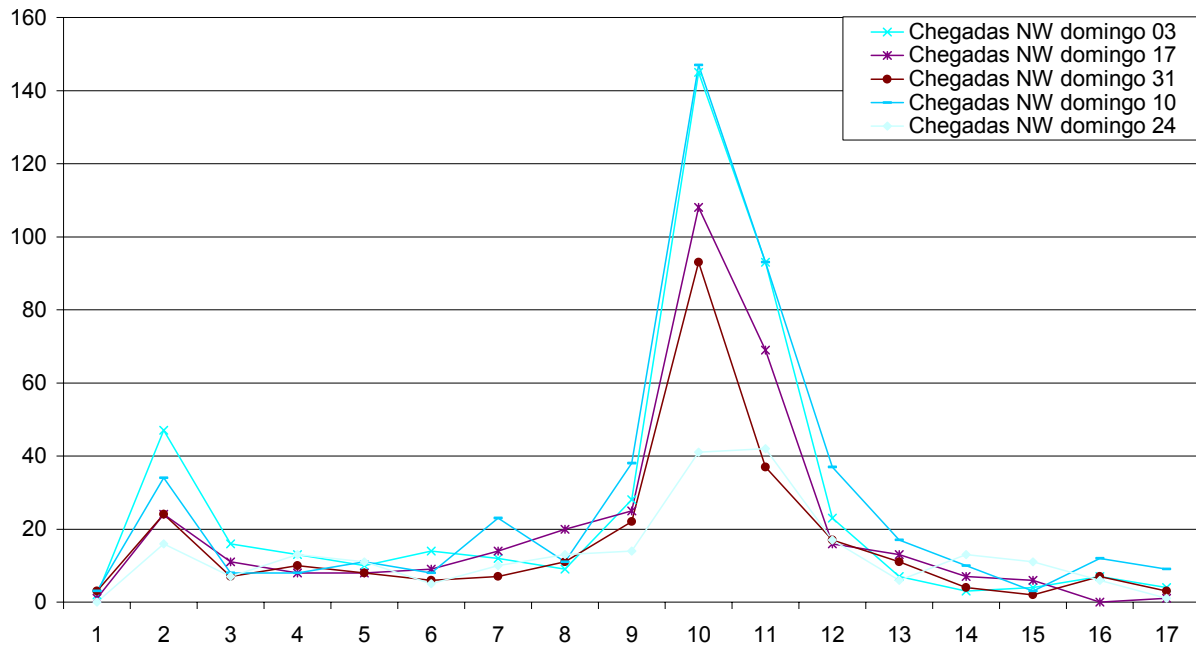


Figura 6: Chegadas do tipo NW aos domingo

O valor médio de chegadas de chamadas dos 17 intervalos de uma hora dos domingos do mês janeiro é tomado como número de chegadas de chamadas para correspondente intervalo do modelo.

4.7. Tempo de serviço

A primeira etapa do serviço se inicia na URA e nesta fase o cliente tem a opção de terminar a ligação ou pedir para falar com um atendente, dando início à segunda etapa do serviço. Desta forma, existem dois tempos de serviço: um na URA, o atendimento eletrônico e outro é o atendimento prestado pelos agentes. Para os tempos de serviço na URA deve-se diferenciar entre tempos de clientes terminando o serviço nesta etapa (auto-atendimento que representam apenas 6,2% dos pedidos de serviço) dos tempos de clientes que pedem para falar com um atendente. O primeiro tipo de cliente leva em média 33,7 segundos e o segundo 8,93 segundos na URA. A Tabela 1 demonstra a distribuição encontrada do tempo de serviço na URA para os clientes que escolhem o serviço dos atendentes.

As ligações que chegam até a segunda etapa do serviço também apresentam tempo de serviço diferenciado de acordo com o tipo de serviço pedido. No outro lado, os atendentes possuem um *mix* de serviços diferenciados e apresentam características diferentes em termos de tempo médio de serviço e desvio padrão. No banco de dados existem vários registros de chamadas com tempo de serviço menor que 10 segundos. Na documentação relacionada ao banco de dados, os autores mencionam que este fenômeno é causado por atendentes desligando chamadas para conseguir mais tempo de descanso. Portanto, na análise, os tempos menores que 10 segundos são desconsiderados. Um pressuposto do modelo em relação ao tempo de serviço dos atendentes é que este tempo pode ser representado por uma única distribuição teórica dependendo do tipo de serviço, ou seja os agentes de determinado tipo de serviço apresentam características iguais. A Tabela 1 demonstra as distribuições teóricas encontradas para cada tipo de chamada.

4.8. Probabilidade de abandonos

O abandono de ligações é um parâmetro importante que tem um impacto significativo no modelo. Neste trabalho, o tempo que o cliente está disposto a ficar na fila até abandonar a ligação é modelado através de uma distribuição teórica ajustada. Os tempos históricos podem ser extraídos diretamente do banco de dados; são os tempos de fila maiores que zero e com resultado “*Outcome*” igual a “*Hang*”. A Tabela 2 demonstra o número de abandonos por tipo de chamada. Pode-se observar que a probabilidade de um abandono depende do tipo de chamada. O tempo de tolerância indica o tempo médio o cliente esta disposto a ficar na fila antes de abandonar a ligação. No modelo este tempo é modelado por uma distribuição única demonstrada na Tabela 1. A probabilidade de um abandono, após este tempo de tolerância, é modelado de acordo com as probabilidades da 3ª coluna, % abandono, da Tabela 2.

Tipo de Chamada	Total Oferecida	Abandonos	% abandono	tempo de tolerância
PS	20114	1650	8%	56.7
PE	133	22	17%	84.73
NW	3937	1204	31%	59.11
TT	715	62	9%	61.29
IN	203	58	29%	134.64
NE	1469	84	6%	120.38

Tabela 2: Características de abandono de ligações no sistema real

4.9. Verificação e Validação do modelo

As seguintes técnicas são empregadas na verificação e validação do modelo:

- ✓ O modelo foi construído em módulos representando os principais mecanismos do processo (as chegadas de chamadas, a fila e os abandonos e o atendimento de chamadas na URA e pelos agentes).
- ✓ A rotina de rastreamento, o Trace, do Arena foi utilizada para o acompanhamento dos eventos ocorridos no modelo.
- ✓ O modelo foi executado com pressupostos simplificados, como número de chegadas por unidade de tempo constante, ausência de abandonos e tempo de serviço, representados por uma distribuição exponencial, para verificar a conformidade com a teoria de filas.
- ✓ Para a parte lógica do processo de encaminhamento de ligações e configuração da Central de Atendimento foi consultado um especialista em centrais telefônicas.

O método de validação empregado aqui é através da comparação dos resultados do sistema real com os resultados do modelo alimentado com dados históricos observados no sistema real, como proposto por Law e Kelton (1991). Neste estudo a maior ênfase é dada neste método de validação. Para os dias 1 até 9 de janeiro, os dados históricos

são transformados num formato adequado para o modelo. O tempo de chegada, tempo de serviço, tempo de abandono do horário de trabalho de cada atendente são modelados conforme os dados reais. O tempo entre o início e fim de duas chamadas consecutivas é modelado como tempo após o encerramento da ligação que o atendente fica indisponível para atender uma nova chamada.

Devido às limitações da versão acadêmica do Arena, apenas as duas primeiras horas de cada dia são modeladas. A versão acadêmica do programa possui uma limitação do número de objetos que o analista pode incluir no modelo. Cada atendente modelado é representado por um módulo que faz a leitura do tempo de pós-ligação a partir de arquivos armazenados no disco do computador. Simulando as primeiras duas horas do dia, limita-se o número de módulos de leitura e evita-se que se ultrapasse o número máximo de objetos.

O critério de desempenho é o tempo médio de fila. Pode-se construir um intervalo de confiança através do teste-t para as diferenças do tempo médio de fila do sistema real e o modelo. A Tabela 5 demonstra a comparação do sistema real com a simulação executada com os dados históricos e o teste-t com um intervalo de confiança de 90%.

Experimento j	Simulação Xj	Dados reais Yj	(Xj – Yj)
1	12.36	11.25	1.112
2	14.53	14.60	-0.068
3	24.15	20.97	3.176
4	8.20	9.06	-0.857
5	26.58	24.77	1.810
6	8.23	6.74	1.487
7	27.14	26.77	0.370
8	6.92	6.73	0.189
9	17.15	17.56	-0.414
<i>Média</i>			0.756
<i>Desvio Padrão</i>			1.262
<i>Semi-Intervalo</i>			2.347
<i>Linf</i>			-1.591
<i>Lsup</i>			3,103

Tabela 3: Teste-t para diferenças do tempo médio de fila

Deve-se observar que o valor zero está contido no intervalo [$L_{inf} = -1,591$, $L_{sup} = 3,103$], portanto qualquer diferença observada entre X_j e Y_j , estatisticamente não é significativa a um nível de 90% e deve ser atribuída a ruído. Na prática, uma diferença tão pequena entre X_j e Y_j não é considerada significativa tendo em vista a característica do tempo de fila de ser altamente sensível à pequenas mudanças nas variáveis do sistema, como por exemplo: tempo de chegada, tempo de serviço número de atendentes etc. Conclui-se que para os fins deste estudo que o modelo é uma boa representação do sistema real.

4.8. O Modelo de Simulação

Como existem horários fixos de início e fim de funcionamento da Central de Atendimento na parte do serviço prestado pelos agentes, o sistema é modelado como

um sistema terminal. A Figura 7 demonstra um fluxograma do processo de ligação e o modelo pode ser resumido da seguinte maneira: Através do processo de geração de chegadas, as entidades (ligações) são lançadas no modelo.

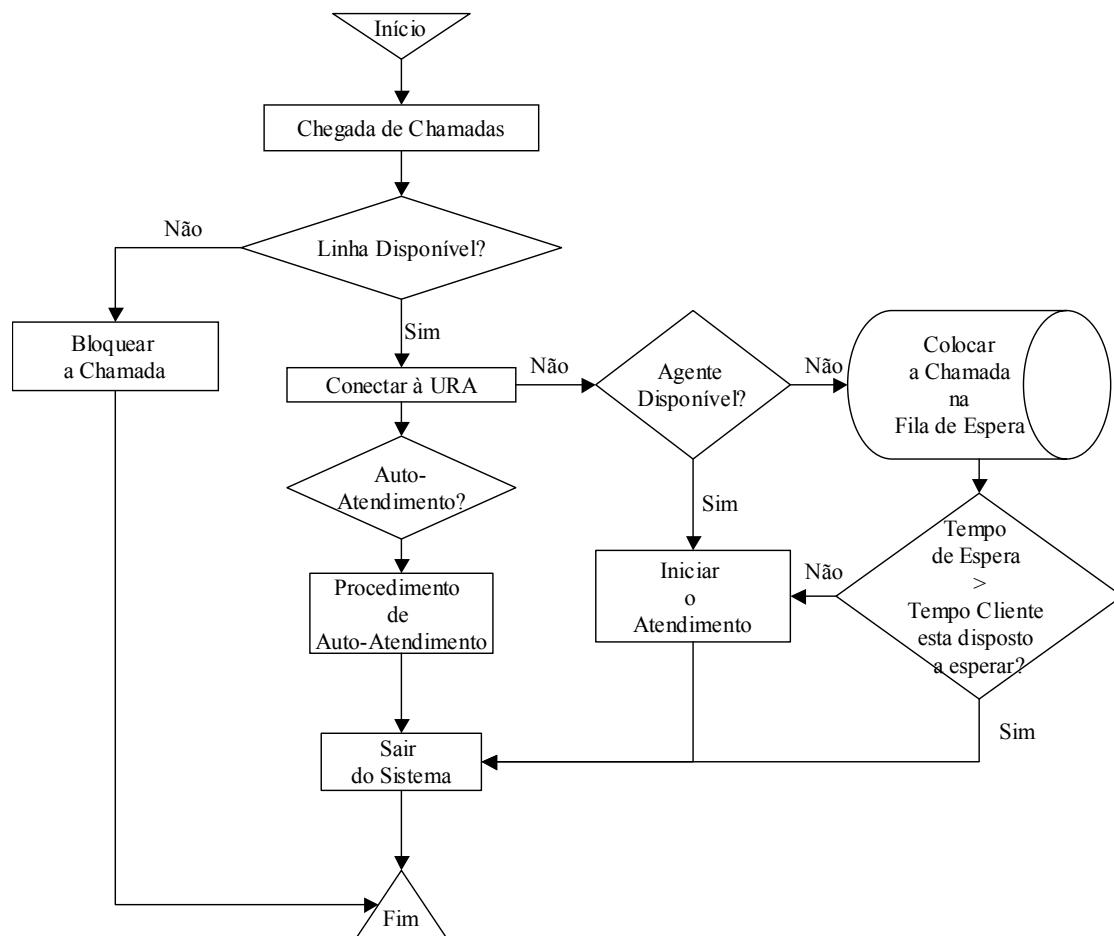


Figura 7: Fluxograma do processo de ligação da Central de Atendimento

Na chegada, se houver uma linha telefônica disponível, modelada como o recurso linhas telefônicas, a linha é tomada pela entidade. Se não encontrar uma linha livre, a entidade é rejeitada do sistema, simulando desta maneira um bloqueio. Depois de ocupar uma linha, a entidade tenta tomar posse de um atendente modelado, como o recurso atendentes. Se tiver um atendente disponível, a entidade é servida com um

tempo de atendimento sorteado da distribuição teórica, representando o tempo de atendimento. Em seguida, a entidade libera o recurso linhas telefônicas e sofre outro atraso representando o tempo de pós-ligação. Depois deste atraso, a entidade libera o recurso atendentes e sai do sistema. Se todos os atendentes estão ocupados a entidade é colocada numa fila de espera aguardando seu atendimento. A modelagem dos abandonos é baseada no tempo de espera acumulada que a entidade gasta na fila de espera; se este tempo ultrapassa um valor sorteado da distribuição representando o tempo que o cliente está disposto a ficar na fila, a entidade é removida da fila simulando assim um abandono.

4.8. Conclusão

Neste capítulo apresentou-se o sistema modelado e o processo de modelagem com auxílio do *software* Arena. Através do *Arena Output Analyzer* as seguintes distribuições representativas foram identificadas: do tempo que o agente fica indisponível para atender chamadas, o tempo de serviço na URA, o tempo de atendimento dos agentes e o tempo que o cliente está disposto a ficar na fila. Demonstrou-se que o modelo do processo de chegadas de chamadas representa, de forma satisfatória, as chegadas, de acordo com o sistema real. O grau de detalhamento obtido no modelo depende dos dados disponíveis ao analista. Por outro lado, a versão acadêmica do *software* limita o grau de detalhamento que se consegue no modelo em relação ao número de objetos que se pode incluir no mesmo. Os processos de verificação e validação demonstraram que o modelo é uma boa representação da Central de Atendimento. A partir de um modelo validado, o analista pode, além de comparar diferentes alternativas de configuração do sistema, também fazer uma série de inferências sobre medidas de desempenho de seu interesse. O capítulo a seguir demonstra experimentação com o modelo.

5. EXPERIMENTAÇÃO

Neste capítulo apresenta-se uma comparação do tempo médio de fila observado no modelo de simulação com o tempo médio de fila calculado pela fórmula de Erlang-C. Em seguida, apresenta-se experimentações analisando o efeito no modelo do momento e a duração do intervalo de trabalho, um aumento no volume de chegadas de chamadas e da mudança da estratégia de roteamento de chamadas. Para fins de comparação de cenários é criado um modelo base representando domingo, o dia mais movimentado no sistema real.

5.1. O modelo base

O modelo base representa o domingo, primeiro dia de trabalho e o dia mais movimentado no sistema. Para cada tipo de chamada o valor médio de chegadas por intervalo de uma hora de todos os domingos de janeiro é tomado como número de chegadas de chamadas do determinado intervalo. A Figura 8 demonstra a média de chamadas de cada tipo de serviço por intervalos de uma hora de duração. A Tabela 1 do capítulo 4 demonstra as distribuições usadas no modelo base. A Tabela 4 apresenta uma estimativa dos horários de trabalho dos atendentes como era no primeiro domingo de janeiro de 1999 e também adotado no modelo base. Os dados são filtrados do banco de dados com auxílio de uma planilha eletrônica e apresentam o início do atendimento da primeira e o fim da última chamada atendida por cada atendente presente naquele domingo.

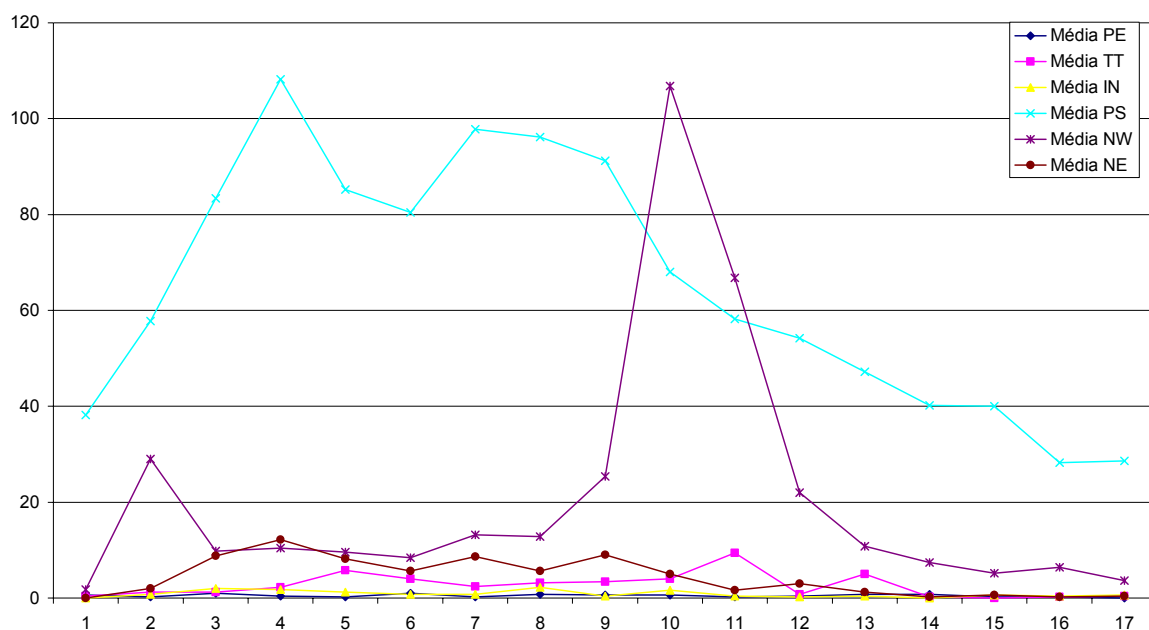


Figura 8: Numero médio de chamadas por período

Atendente	Início do horário do atendente	Fim do horário do atendente
GILI	06:54:50	13:32:15
TOVA	06:58:31	14:51:48
LORI	07:00:14	13:58:57
TALI	07:10:37	13:46:03
SHARON	07:48:32	16:00:47
YITZ	07:49:50	16:01:57
KAZAV	08:00:55	15:55:09
ZOHARI	08:28:20	15:06:21
ELI	08:45:15	15:09:47
AVNI	09:01:17	17:00:40
YIFAT	09:06:42	16:58:12
PINHAS	11:01:20	16:59:39
MICHAL	12:54:53	20:39:18
ANAT	13:59:28	20:55:51
MIKI	13:59:43	22:03:19
NAAMA	14:01:44	21:43:09
IDIT	15:04:17	21:58:35
MORIAH	15:55:14	23:52:25
GELBER	16:22:21	00:00:25
DORIT	16:56:49	23:00:28
BENSION	17:58:56	00:01:09
SHLOMO	19:23:32	00:00:25

Tabela 4: Horário dos atendentes

No sistema real as chamadas são roteadas a grupos de atendentes. O Quadro 1 demonstra o esquema de roteamento de chamadas. Pode-se observar que o mesmo atendente pode pertencer a mais de um grupo.

Atendente	Tipo de Chamada					
	PS	PE	NW	TT	IN	NE
ANAT	1		1	1	1	1
AVNI	1		1	1		1
BENSION	1		1	1	1	1
DORIT	1	1	1	1		
ELI	1		1	1		1
GELBER	1		1	1		
GILI	1	1		1		1
IDIT	1	1	1	1		
KAZAV	1		1	1		1
LORI	1		1	1	1	
MICHAL	1	1	1	1	1	
MIKI	1	1	1	1		
MORIAH	1	1	1	1		1
NAAMA	1		1	1		
PINHAS	1		1	1		
SHARON	1		1	1	1	1
SHLOMO	1		1	1	1	
TALI	1	1	1	1		
TOVA	1	1		1	1	1
YIFAT	1	1	1	1	1	1
YITZ	1	1	1	1	1	1
ZOHARI	1		1	1		1

Quadro 1: Esquema de roteamento de chamadas

5.2. Comparação de Erlang-C com a Simulação

Apesar da disponibilidade de programas computacionais de simulação voltados às centrais de atendimento na prática muitos gerentes usam calculadores de Erlang-C na configuração da sua central de atendimento. Pode-se fazer uma comparação entre os dois métodos de análise. As suposições necessárias para aplicar as fórmulas de Erlang-C são discutidas no capítulo 3. O tempo médio de serviço tomado para todos os tipos de chamadas oferecidas é 197 segundos, somando o tempo médio de pós-ligação de 54 segundos fornece o tempo médio de serviço 251 segundos, necessário para o cálculo do tempo de fila com a fórmula de Erlang-C. O sistema real apresenta um tempo médio de fila de 35 segundos. Toma-se este tempo como tempo médio de fila permitido e se for estabelecido que este tempo deve ser atingido 80% do tempo, pode-se calcular o tempo de espera por intervalo de uma hora para determinado número de atendentes presentes. Outra suposição que serve igualmente para a simulação como para a fórmula de Erlang-C é que não há tempo de intervalo para os atendentes. A Tabela 5 demonstra o tempo de fila calculado pela fórmula de Erlang-C e o tempo de fila médio observado na simulação com 50 replicações. Por efeito de comparação a 4ª coluna demonstra o tempo médio observado no sistema real do domingo. A Figura 9 demonstra o gráfico do tempo de fila de cada período como calculado ou observado. Pode-se observar que a fórmula de Erlang-C está superestimando o tempo de fila nos períodos com um alto o número de chegadas. Nos demais períodos os cálculos são razoavelmente próximo à simulação. Deve-se observar que o tempo de fila médio no sistema real, excluindo os períodos 1, 10 e 11 é sempre maior do que na simulação isto devido aos prováveis intervalos de trabalho tomados pelos atendentes no sistema real. Para os períodos 1, 10 e 11 a simulação apresenta maior tempo de fila que deve ser atribuído aos efeitos transientes do modelo que depende do momento e duração do intervalo como será mostrado na seção a seguir.

Período	Número de Atendentes	Número de Chegadas na Simulação e usado para o Erlang-C	Tempo médio de Fila Erlang C (segundos)	Tempo médio de Fila Simulação (segundos)	Tempo médio de Fila Sistema Real (segundos)
1	6	41	7	49	17
2	9	91	24	24	21
3	11	105	11	6	27
4	11	136	90	28	40
5	12	115	9	7	27
6	12	103	4	2	20
7	13	125	8	8	32
8	12	122	14	3	29
9	11	133	72	21	32
10	11	188	>209	97	65
11	9	146	>225	83	54
12	8	81	30	8	21
13	9	70	4	1	9
14	9	49	0	0	7
15	8	46	1	2	11
16	4	35	48	6	18
17	4	33	38	19	17

Tabela 5: Tempo médio de fila através de Erlang-C e as observações no modelo

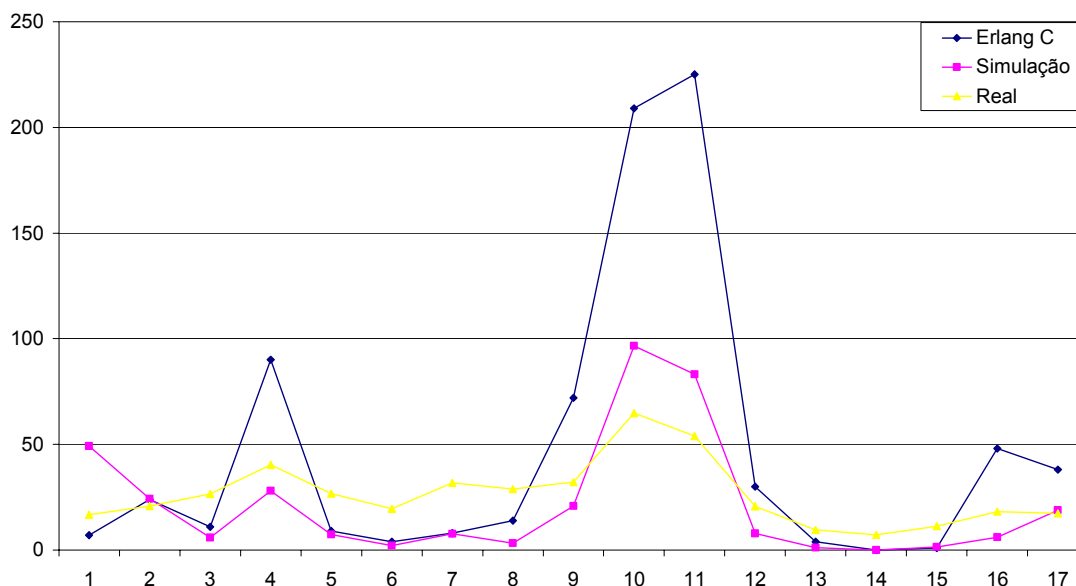


Figura 9: Comparação de Erlang-C com as observações do modelo e sistema real

5.3. Análise em relação ao número de atendentes

Pode-se analisar o efeito da saída de um atendente no tempo de fila do modelo. Se por exemplo um atendente chegar 15 minutos atrasado ou tomar neste momento seu intervalo o tempo de fila principalmente no 10º período, o período de maior movimento, se altera. A primeira parte do Quadro 2 demonstra o teste-t para diferenças do tempo médio de fila do 10º período das duas alternativas: o modelo base e o modelo com um agente iniciando o atendimento 15 minutos mais tarde. O teste-t indica que o tempo médio de fila é 11,5 segundos menor em relação ao modelo base. Uma análise mais precisa revela que o momento de um intervalo pode provocar um aumento ou uma diminuição no tempo de fila nos períodos posteriores. No caso anterior houve uma diminuição no tempo médio de fila. Entretanto, se o intervalo de 15 minutos acontece 150 minutos depois do início do serviço do atendente, o tempo médio de fila aumenta com 20 segundos no 10º período e 27 segundos no 11º período como mostra o Quadro 1. Se a duração do intervalo for aumentado para 20 minutos o tempo médio de fila no 10º período diminui e com um intervalo de 30 minutos ele decresce ainda mais. A

Figura 10 demonstra graficamente as alterações no tempo médio de fila para as diferentes situações.

Paired-T Means Comparison :						
IDENTIFIER	ESTD. MEAN DIFFERENCE	STANDARD DEVIATION	0.950 C.I. HALF-WIDTH	MINIMUM VALUE	MAXIMUM VALUE	NUMBER OF OBS
PERIOD_10	11.5	135	3.02	0	644	7715
				p	575	7715
REJECT H0 => MEANS ARE NOT EQUAL AT 0.05 LEVEL						
Paired-T Means Comparison :						
IDENTIFIER	ESTD. MEAN DIFFERENCE	STANDARD DEVIATION	0.950 C.I. HALF-WIDTH	MINIMUM VALUE	MAXIMUM VALUE	NUMBER OF OBS
PERIOD_10	-20.6	158	3.53	0	644	7715
				0	827	7715
REJECT H0 => MEANS ARE NOT EQUAL AT 0.05 LEVEL						
Paired-T Means Comparison :						
IDENTIFIER	ESTD. MEAN DIFFERENCE	STANDARD DEVIATION	0.950 C.I. HALF-WIDTH	MINIMUM VALUE	MAXIMUM VALUE	NUMBER OF OBS
PERIOD_11	-27.2	186	4.72	0	724	5969
				0	826	5969
REJECT H0 => MEANS ARE NOT EQUAL AT 0.05 LEVEL						

Quadro 2: Teste-t para a o tempo médio de fila das alternativas comparadas

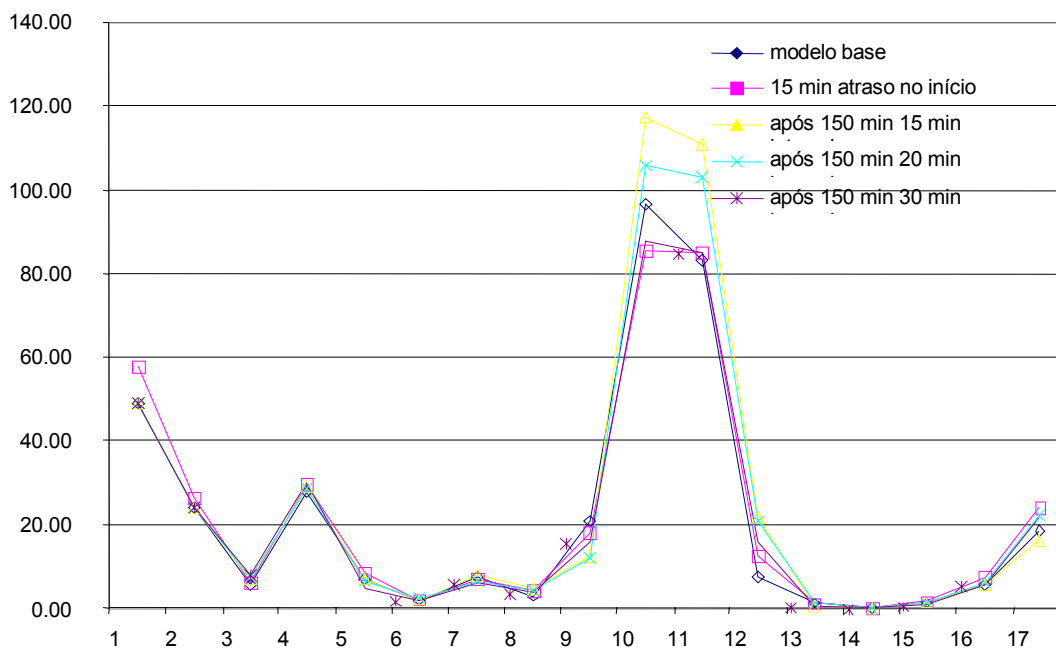


Figura 10: Gráfico do tempo médio de fila para 4 diferentes intervalos de trabalho

5.4. Análise em relação ao volume de chamadas

As chamadas do tipo PS e NW representam a maior proporção de chamadas direcionadas aos atendentes. Primeiramente o número de chamadas do tipo PS é aumentado com 10% depois as chamadas do tipo NW também com 10%. Finalmente o volume de chamadas dos dois tipos ao mesmo tempo, também com 10% cada um. A Figura demonstra graficamente o resultado dos aumentos no volume de chegadas no tempo médio na fila de espera. Em todos os casos existe um aumento considerável da média no período de maior movimentação enquanto nos demais intervalos de tempo as alterações não são significantes. O pico é causado pelo aumento de chegadas das chamadas do tipo NW. Segundo a documentação as chamadas do tipo NW representam ligações de clientes novos. O pico às 16 até 18 horas, entretanto, sugere que existe algum acontecimento neste período que provoca este aumento no número

de chegadas do tipo NW. Uma explicação possível é que o tipo de chamadas NE e NW são trocadas na documentação. Segundo a documentação, NE representa chamadas envolvendo atividades na bolsa de valores e é mais provável que este tipo de chamada apresentar um pico no horário do fechamento da bolsa. Pode-se concluir que alterações no volume de chamadas, a duração e momento de intervalo de descanso apresentam o maior efeito no período de maior movimentação.

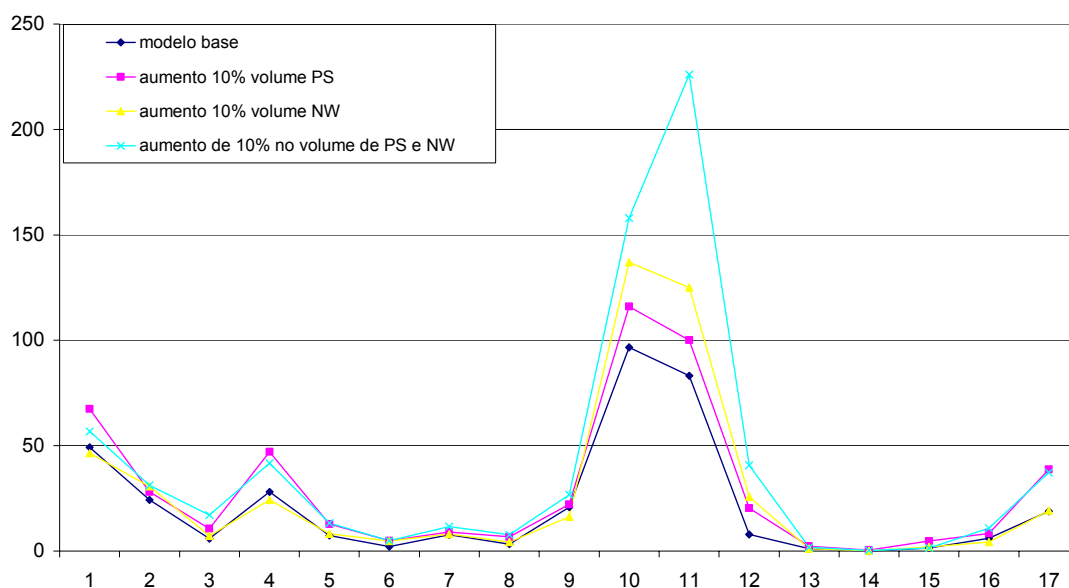


Figura 11: O efeito do aumento do número de chegadas no tempo médio de fila

5.5. Atendentes especializadas versus atendentes multifuncionais

O Quadro 1 demonstra o tipo de chamada atendida pelos 22 agentes individuais no sistema real que fazem parte do modelo básico. Pode-se observar que as chamadas do tipo PS e TT são roteadas para todos os atendentes, enquanto as demais chamadas são roteadas apenas para grupos selecionados. Na terminologia de Centrais de Atendimento essa estratégia é chamada de *skills-based routing*. Aplicando o *skills-based routing* na prática torna a programação de turnos mais complexa e, quando só

existem grupos especializados, o número de chamadas por servidor se torna menor e a utilização dos servidores diminui.

Pode-se comparar a estratégia atual empregada na Central de Atendimento com uma estratégia onde os atendentes são multifuncionais tomando como critério de desempenho o tempo de espera na fila. Em termos operacionais, a alternativa com o menor tempo médio de fila é considerada a melhor. A Figura 12 demonstra o gráfico do tempo médio de fila em cada intervalo de uma hora para o modelo base e o modelo representando a estratégia empregando atendentes multifuncionais. Pode-se observar que não existe uma diferença significativa entre as duas alternativas. Isto porque as chamadas PS e NW, que apresentam a maior proporção de chamadas oferecidas, já estão sendo atendidas na estratégia atual por praticamente todos os agentes. Operacionalmente uma estratégia de atendentes multifuncionais não melhora o tempo médio de fila e não há razões investir em treinamento para transformar os atendentes especializadas em atendentes multifuncionais.

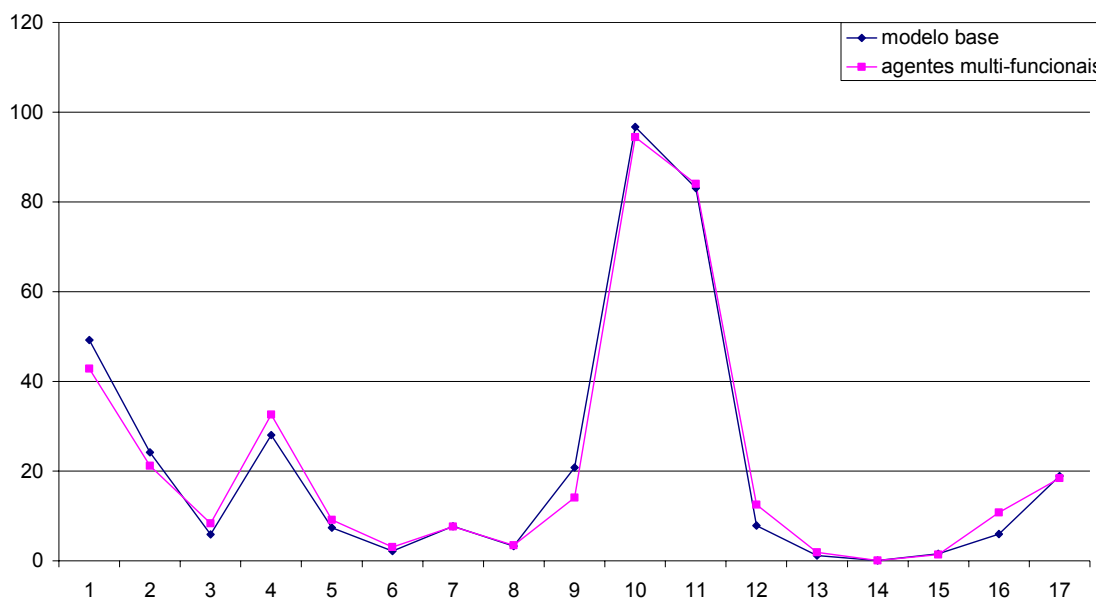


Figura 12: Tempo médio de fila do modelo base e o modelo alternativo

5.5. Conclusão

Neste capítulo apresentou-se uma série de experimentos com o modelo. O tempo médio de fila por intervalo de tempo de uma hora observado no modelo base foi comparado com o cálculo do tempo médio de fila através da fórmula de Erlang-C. Foi observado que a fórmula de Erlang-C está superestimando o tempo médio de fila quando ocorre um pico no número de chegadas num determinado período. Nos demais períodos o cálculo do tempo médio de espera pela fórmula de Erlang-C é próximo ao valor observado no modelo. O número de atendentes presentes tem um grande efeito no tempo médio de fila. A saída de um atendente para intervalo de trabalho, caracterizado pelo momento e duração da saída, tem seu maior efeito no período de maior movimentação. Existe uma relação entre o tempo médio de fila, o momento e a duração do intervalo. Para os casos analisados, o tempo médio de fila no período de maior movimentação diminui quando o tempo do intervalo é maior. Aumentar o volume de chegadas de chamadas também tem seu maior efeito no tempo de fila médio durante o tempo de maior movimentação. No capítulo seguinte, apresentam-se as conclusões que puderam ser feitas em função da elaboração deste estudo de simulação.

6. CONCLUSÕES E SUGESTÕES

6.1. Conclusões

A metodologia da modelagem e simulação proposta pela literatura foi aplicada em um estudo de caso de uma central de atendimento por telefone. Pela a aplicação dos métodos propostos foi possível obter um melhor entendimento da análise e os passos envolvidos em um estudo de simulação. Como maior dificuldade encontrada se destaca a falta de acesso aos gerentes da central para verificar as várias suposições feitas na modelagem da central e a possibilidade de obter dados mais detalhados. Em função dos experimentos conduzidos pode-se concluir que o número de chegadas de pedidos de serviço, o momento e duração dos intervalos de trabalho dos atendentes provocam um impacto significativo no critério de desempenho escolhido.

Ao longo deste trabalho foram apresentados e discutidos os principais conceitos e técnicas envolvidos na modelagem e programas de computadores voltados à simulação de sistemas. No capítulo 2 foi apresentado como a dinâmica e aleatoriedade dos sistemas reais são incorporadas aos programas de simulação. No mesmo capítulo, tratou-se da modelagem dos dados de entrada para a alimentação do modelo de simulação. O número mínimo de replicações necessário para alcançar determinada precisão nos resultados também foi assunto deste capítulo. Na parte final do mesmo foram apresentados os passos de um projeto de simulação, com ênfase na etapa de verificação e validação do modelo que deve garantir que o modelo apresenta comportamento semelhante ao sistema modelado.

No capítulo 3 apresentou-se a definição de uma Central de Atendimento e uma revisão da literatura relacionada às Centrais de Atendimento. Demonstrou-se que existem várias limitações na aplicação dos métodos analíticos tradicionais na configuração de uma Central de Atendimento em termos de número de recursos necessários para atender a demanda de serviço. A maioria dos artigos revisados sugere a aplicação de simulação na análise de uma Central de Atendimento. No final do

capítulo foram apresentadas medidas de desempenho empregadas nas Centrais de Atendimento.

No capítulo 4 tratou-se da modelagem da Central de Atendimento do Banco Anônimo com o auxílio do programa computacional Arena 3.0. Pode-se concluir que o grau de detalhamento que se consegue incluir num modelo depende dos dados disponíveis ao analista. Análise mais precisa do banco de dados sugere a existência de um tempo pós-ligação que o agente fica indisponível para atender novas chamadas. Com os dados disponíveis, entretanto, não foi possível filtrar este tempo de pós-ligação do banco de dados ao menos a duração de intervalos de trabalho de cada atendente. O modelo construído é uma boa representação do sistema real mediante o teste estatístico tal como foi visto no processo de validação. A inclusão do tempo entre o fim e início de duas chamadas consecutivas foi fundamental na validação do modelo. No processo de validação, o modelo foi executado com dados históricos do sistema real através da leitura de arquivos com estes dados armazenados no computador. O método de “rejeição” utilizado na geração de chegadas de chamadas representa um modelo razoável de um processo de Poisson não estacionário. No final do capítulo 4 foi apresentado um resumo sucinto do modelo de simulação como modelado na ambiente do Arena 3.0.

No capítulo 5 apresentou-se uma série de experimentos com o modelo. O método analítico utilizando a fórmula de Erlang-C foi comparado com resultados da simulação. Foi observado que a fórmula de Erlang-C está superestimando o tempo médio de fila nos períodos que apresentam um pico no número de chegadas de chamadas. Nos demais períodos o cálculo do tempo médio de espera pela fórmula de Erlang-C fornece uma estimação razoavelmente próximo aos observações no modelo. O número de atendentes presentes tem um grande impacto no tempo médio de fila. A saída de um atendente para intervalo de trabalho caracterizado pelo momento e duração da saída tem seu maior efeito no período de maior movimentação. Existe uma relação entre o tempo médio de fila o momento e a duração do intervalo. Para os casos analisados, o tempo médio de fila no período de maior movimentação diminui quando o tempo do intervalo é maior. Aumentar o volume de chegadas de chamadas também tem seu maior efeito no tempo de fila médio durante o tempo de maior movimentação quando

ocorre o maior número de chegadas de chamadas. Em termos operacionais a estratégia de roteamento de chamadas para todos os atendentes não apresentou diferença significativa comparada com o *skills-based routing*.

6.2. Recomendações e Sugestões

Recomenda-se aos gerentes da central de atendimento as seguintes ações:

Usar o modelo de simulação como ferramenta de previsão para avaliar melhoramentos no nível de serviço, para um dado nível de recursos principalmente durante o período de maior movimentação

Analisar a natureza das chamadas do tipo NW responsável pelo pico com o objetivo de influenciar o comportamento do cliente para evitar este pico sem aumentar o número de atendentes.

Padronizar o atendimento para diminuir o tempo médio e a variância do serviço dos atendentes através de treinamento principalmente dos serviços de maior proporção.

Recomenda-se para trabalhos futuros as seguintes ações:

Analisar a relação que existe entre o tempo médio de fila no período de maior movimentação, o momento e duração do intervalo de trabalho para testar se a escolha do momento e duração do intervalo pode diminuir o tempo de fila no período de maior movimentação.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ANDREWS, Bruce; PARSONS, Henry. **Establishing Telephone-Agent Staffing Levels through Economic Optimatization**. Interfaces, Vol. 23, No. 2, março-abril 1993.

CHOKSHI, Rupesh. **Decision Support for Call Center Management Using Simulation**. In: Proceedings of the 1999 Winter Simulation Conference, Farrington, P. A. et al., p. 1634-1637.

COOPER, R. B. **Introduction to Queueing Theory**. 2 ed. New York: North-Holland, 1981.

DIAGNOSTIC STRATEGIES. **Traffic Modelling an Resource Allocation in Call Centers**. Needham, 2000. Disponível em < www.diagnosticstrategies.com >. Acesso em: 11 nov. 2002.

DUDER, John C.; ROSENWEIN, Moshe B. **Towards “zero abandments “ in Call Center performance**. European Journal of Operational Research, Vol. 135, 2001. p. 50-56.

FREITAS FILHO, Paulo José de. **Introdução à Modelagem e Simulação de Sistemas-com Aplicações em Arena**. Florianópolis: Visual Books, 2001.

GANS, Noah; KOOLE, Ger; MANDELBAUM, Avishai. **Telephone Call Centers: Tutorial, Review, and Research Prospects**. Technion, Out . 2002. Disponível em <http://ie.technion.ac.il/~serveng>. Acesso em: 19 dez. 2002.

HARRIS, Carl M.; HOFFMAN Karla L. **Modelling the IRS Telephone Taxpayer Information System**. Operations Research, Vol. 35, No 4, p. 504-523, Julho-Agosto 1987.

KELTON, W. David; SADOWSKI, Randall P; SADOWSKI, Deborah A. **Simulation with Arena**. 1ª ed. Boston: McGraw-Hill, 1998.

KELTON, W. David; SADOWSKI, Randall P; SADOWSKI, Deborah A. **Simulation with Arena**. 2ª ed. Boston: McGraw-Hill, 2002.

KLUNGE, Roger. **Simulation of a Claims Call Center: A Success and a Failure.** In: Proceedings of the 1999 Winter Simulation Conference, Farrington, P. A. *et al.*, p. 1648-1653.

KOOLE, Ger; MANDELBAUM, Avishai. **Queueing Models of Call Centers: An Introduction.** Technion, Out . 2001. Disponível em <http://ie.technion.ac.il/~serveng>. Acesso em: 4 dez. 2002.

KOOLE, Ger. **Call Center Mathematics: A Scientific Method for understanding and improving your Contact Center.** Amsterdam, 2001. Disponível em <www.cs.vu.nl/~koole/ccmath>. Acesso em: 16 dez. 2002.

LAW, Averill M; KELTON, W. David. **Simulation Modelling and Analysis.** 2 ed. New York: McGraw-Hill, 1991.

SANGARAYAKUL, Bordin. **Use of Simulation to Analyze Block Manufacturing Methods.** 1998. Dissertação (Mestrado em Engenharia Civil) – Virginia Polytechnic Institute and University, Blacksburg, Virginia.

SHANNON, Robert E. **Introduction to the Art and Science of Simulation.** In: Proceedings of the 1998 Winter Simulation Conference, Medeiros D.J. *et al.*, p. 7-14.

TANIR, Oryal; BOOTH, Richard. **Call Center Simulation in Bell Canada.** In: Proceedings of the 1999 Winter Simulation Conference, Farrington, P. A. *et al.*, p. 1640-1647

ANEXOS

AMOSTRA DO BANCO DE DADOS

Uma amostra do Banco de Dados disponível na internet, pagina: <http://ie.technion.ac.il/~serveng>

URA + linha	Id. Cham.	Cliente	prioridade	tipo	data	Entrada URA	Saída URA	Tempo URA	Início fila	Fim fila	Tempo fila	resultado	Início serviço	Fim serviço	Tempo serviço	servidor
AA0106	34054	22090179	2	PS	990101	0:00:02	0:00:08	6	0:00:08	0:02:28	140	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0101	33116	9664491	2	PS	990101	0:00:31	0:00:36	5	0:00:36	0:03:09	153	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0101	33117	0	0	PS	990101	0:34:12	0:34:23	11	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0102	28423	0	0	PS	990101	0:36:59	0:36:59	0	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0103	24036	0	0	PS	990101	0:37:32	0:37:32	0	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0104	20696	0	0	NW	990101	0:39:43	0:40:17	34	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0105	33290	0	0	PS	990101	0:45:42	0:46:16	34	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0106	34055	0	0	PS	990101	0:48:35	0:49:05	30	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0109	27869	0	0	PS	990101	0:48:36	0:48:54	18	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0110	27484	0	0	PS	990101	0:50:56	0:50:56	0	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0115	12077	0	0	NW	990101	0:51:23	0:51:58	35	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0201	33333	0	0	PS	990101	1:16:58	1:17:34	36	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0202	32426	0	0	NW	990101	1:19:19	1:19:45	26	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0203	33140	0	0	NW	990101	1:20:06	1:21:20	74	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0204	32025	0	0	PS	990101	2:18:11	2:18:48	37	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0205	33509	0	0	NW	990101	3:44:57	3:45:20	23	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0206	34175	0	0	PS	990101	3:50:11	3:50:28	17	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0209	30997	0	0	PS	990101	5:10:33	5:11:27	54	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0210	30715	0	0	PS	990101	6:08:06	6:08:19	13	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0213	11996	0	0	PS	990101	6:41:18	6:41:18	0	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0214	12065	0	0	AA	990101	6:41:36	6:41:47	11	0:00:00	0:00:00	0	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0215	11623	52928306	2	PS	990101	6:42:47	6:42:53	6	6:42:53	6:43:26	33	HANG	0:00:00	0:00:00	0	NO_SERVER
AA0101	33118	27997683	2	PS	990101	6:55:20	6:55:26	6	6:55:26	6:55:43	17	AGENT	6:55:43	6:56:37	54	MICHAL
AA0107	2023	0	0	TT	990101	7:04:42	7:04:44	2	0:00:00	0:00:00	0	AGENT	7:04:43	7:04:43	0	YITZ
AA0108	1169	31759814	1	TT	990101	7:05:06	7:05:09	3	0:00:00	0:00:00	0	AGENT	7:05:08	7:05:08	0	BASCH
AA0102	28424	0	0	PS	990101	7:06:32	7:06:42	10	0:00:00	0:00:00	0	AGENT	7:06:41	7:08:28	107	MORIAH
AA0103	24037	0	0	PS	990101	7:08:58	7:09:07	9	0:00:00	0:00:00	0	AGENT	7:09:07	7:11:07	120	MICHAL

O Modelo da Central de Atendimento em Arena 3.0

Dados do modelo

Os dados necessários para o modelo são incorporados da seguinte maneira: O número de chegadas observado por período é representado pelo o vetor *PEDIDOS* com dimensão igual a 34 (34 períodos de meia hora considerando uma expedição de 17 horas por dia) modelado através do módulo *VARIABLES*. Os atendentes (22 em total) são representados pelo módulo *RESCOURCES*. Cada um com sua própria jornada de trabalho.

O módulo *SET* define os grupos de atendentes com habilidades iguais.

As linhas telefônicas (65 linhas) e a URA (32 unidades) são representadas pelos módulos *RESCOURCES*.

As distribuições teóricas utilizadas no modelo mostradas na Tabela 1 são representadas pelo módulo *EXPRESSIONS*.

O tempo simulado é 61.200 segundos por dia

A Figura 1 demonstra os módulos contendo os dados do modelo.

Distribuição	Expressão no Arena
Tempo de Serviço na URA	$10 + \text{LOGN}(4.17, 3.87)$
Tempo de Tolerância	$0.999 + \text{ERLA}(60.1, 1)$
Tempo Pós-Ligação	Empírica
Tempo de Serviço PS	$10 + \text{LOGN}(193, 243)$
Tempo de Serviço PE	$10 + \text{GAMM}(206, 1.31)$
Tempo de Serviço NE	$10 + \text{LOGN}(359, 550)$
Tempo de Serviço NW	$10 + \text{LOGN}(108, 175)$
Tempo de Serviço TT	$10 + \text{WEIB}(127, 0.729)$
Tempo de Serviço IN	$10 + \text{LOGN}(282, 397)$

Tabela 1: Distribuições do modelo

Dados

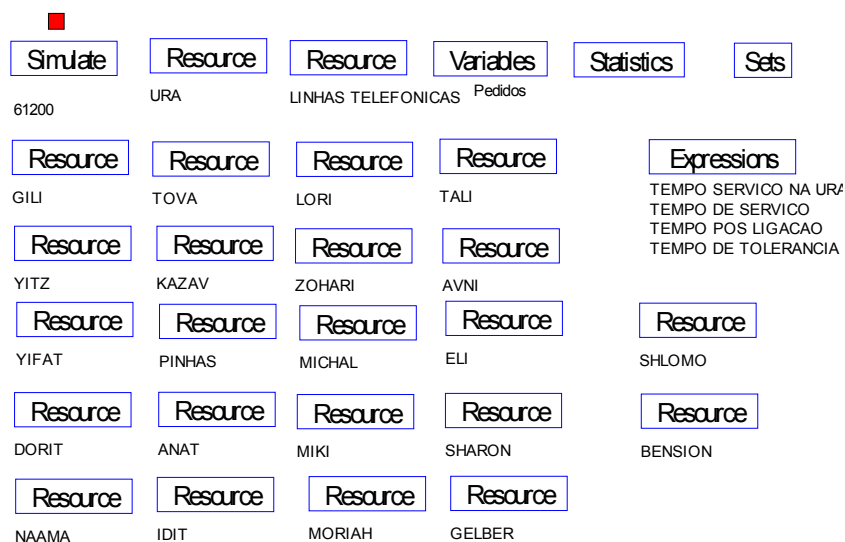


Figura 1: Módulos de Dados de Entrada do modelo

Modelo de Chegadas de Chamadas

O processo de chegadas de chamadas é representado por um processo de Poisson não-estacionário. Segundo Kelton et al., (1998) uma maneira óbvia porem, incorreta seria modelar o tempo entre chegadas do processo por uma distribuição exponencial com uma média representada por uma variável e alterar seu valor de acordo com o número de chegadas observado em cada período. Essa abordagem cria problemas quando a função de número de chegadas por unidade de tempo apresenta mudanças bruscas. Por exemplo, na mudança de um período, onde o número de chegadas é pequeno, para o próximo período, onde existe um grande número de chegadas, o tempo entre chegadas na última fração de tempo no primeiro período é gerado de acordo com um tempo médio entre chegadas relativamente grande, ou seja um pequeno número de chegadas. Isto resulta também, num pequeno número de

chegadas no próximo período, enquanto o número de chegadas deveria ser grande neste intervalo.

Para evitar este problema o método chamado de “rejeição” (Lewis e Shedler, 1979) é aplicado: primeiramente o dia é dividido em intervalos de tempo iguais, neste estudo períodos de 30 minutos. Determina-se em seguida o valor máximo observado do número de chegadas por unidade de tempo λ_{max} de todos os períodos. Um processo de Poisson estacionário é gerado utilizando λ_{max} como número médio de chegadas para todos os intervalos. Cada chegada gerada, entretanto, é rejeitada com uma probabilidade. A probabilidade de uma chegada gerada pelo o processo de Poisson estacionário entrar no sistema durante o período i é igual ao número de chegadas observado no período i , λ_i , dividido pelo valor máximo de chegadas λ_{max} . Isto gera uma função de número de chegadas por unidade de tempo razoavelmente próximo á realidade.

O modelo de chegadas possui três mecanismos propostos pelo Kelton et al., (1998): indicador do período atual, um mecanismo para encontrar λ_{max} , e um mecanismo para criar e lançar as chegadas no modelo.

Indicador do período atual

Essa lógica como mostrada na Figura 2 serve para indicar o período atual. Através de um módulo *CREATE* uma entidade de controle é criada iniciando o dia e para incrementar a variável *Período* depois um atraso igual ao intervalo de tempo de 30 minutos. Quando o número de períodos igual a 34 (o total número de intervalos de 30 minutos por dia) a entidade de controle sai do modelo.

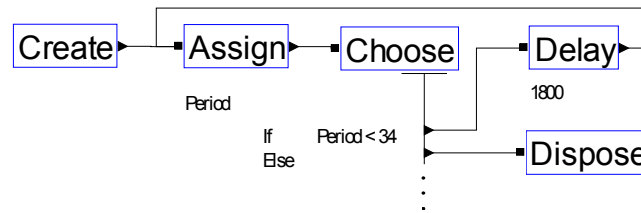


Figura 2: Incrementar o Período

Mecanismo para encontrar λ_{max}

A Figura 3 demonstra a lógica usada para encontrar λ_{max} , com auxílio do modulo *FINDJ* que serve para encontrar uma variável que satisfaz determinada condição neste caso o valor máximo do vetor *PEDIDOS*. O módulo *ASSIGN* atribuí este valor máximo à variável *MAX RATE*.

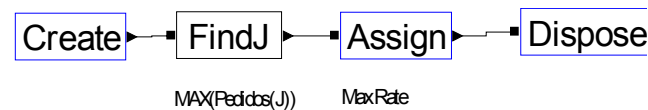


Figura 3: Encontrar λ_{max}

Mecanismo para criar e lançar as chegadas no modelo

Essa lógica representa o método de “rejeição” descrito acima. A lógica é mostrada na Figura 4. O tempo entre chegadas é representado por uma distribuição exponencial com média igual a $3600/\lambda_{max}$ segundos. Cada chegada é lançada no modelo com uma probabilidade igual a $PEDIDOS(PERODO)/\lambda_{max}$.

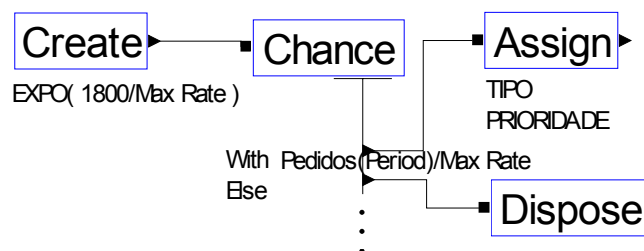


Figura 4: Geração de chegadas

O processo de Serviço

Depois a sua criação a entidade entra no módulo *ASSIGN* que determina o tipo de chamada de acordo com as proporções demonstrada na Tabela 2. O mesmo módulo é usado para determinar o tipo de cliente em relação a preferência usada pelo Banco. Depois isso a entidade tenta tomar posse do recurso Linhas telefônicas que tem uma fila de espera com capacidade zero. Isto significa que quando não existe uma linha disponível a entidade (chamada) é bloqueada a prosseguir, sai do modelo e é contada por um contador como um bloqueio. As entidades que conseguem ocupar uma linha a seguir tomam posse de uma unidade da URA. A seguir ocorre o atraso que representa o serviço na URA representado pela expressão Tempo de Serviço na URA. Depois este atraso a entidade libera o recurso URA. A seguir a entidade encontra um módulo *ASSIGN* que atribuí um número identificador a cada entidade (veja a seção: O Processo de Abandono de Ligações) e o tempo de tolerância do cliente através da expressão Tempo de Abandono. O módulo *DUPLICATE* cria uma entidade igual ao original para o processo de abandono. A entidade original prossegue e a combinação dos módulos *CHOOSE* e *ASSIGN* faz com que 1,5 minutos do horário de chegada dos cliente preferências seja subtraída para avançar este tipo de cliente na fila de espera como acontece no sistema real. O módulo *CHOOSE* identifica o tipo de serviço pedido fazendo com que a entidade entra

numa fila de espera se for necessária e tomar posse do servidor de acordo com o tipo de serviço pedido. A entidade avança no modelo depois o atraso representando o tempo de serviço representado pela expressão Tempo de Serviço(Tipo). Depois o atraso a entidade libera o recurso Linhas telefônicas e acontece um novo atraso que representa o tempo pós-ligação o agente precisa para completar o atendimento. Este tempo é representado pela expressão Tempo Pós-Ligação. A seguir a entidade libera o recurso Atendentes e depois a contagem do número de chamadas atendidas sai do modelo. A Figura 5 demonstra a lógica do processo de Serviço.

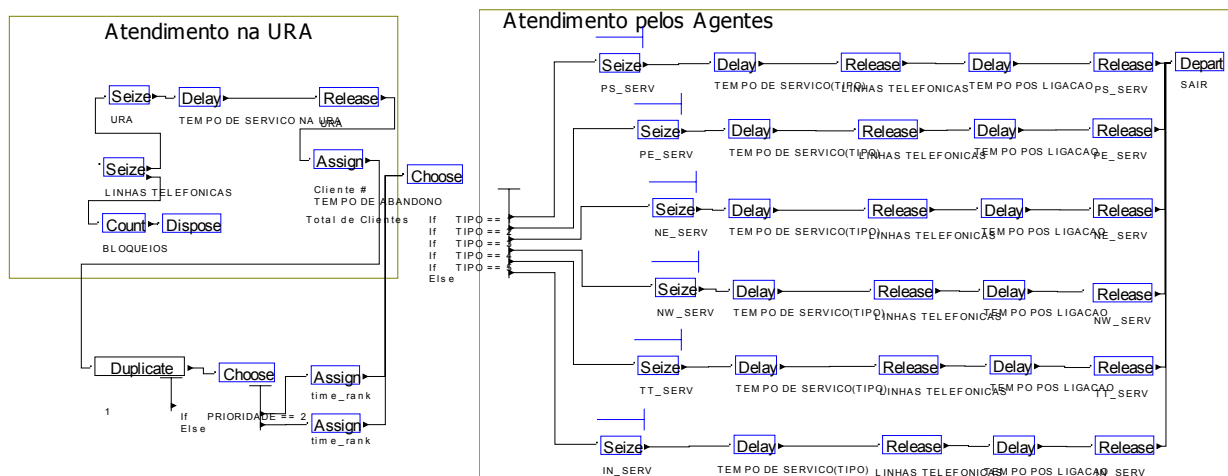


Figura 5: Atendimento de Chamadas na URA e pelos Agentes

O Processo de Abandono de Ligações

O módulo *ASSIGN* que se encontra antes do módulo *DUPLICATE* atribui a cada entidade um número identificador isto serve para modelar o processo de abandono de ligações. A entidade duplicada funciona como uma entidade de controle. Após de sofrer um atraso que representa o tempo de tolerância o

cliente esta disposto a ficar na fila de espera a entidade entra no módulo *CHANCE* que determina a probabilidade de um suposto abandono. A seguir é determinada uma variável *PROCURAR* igual ao número identificador anteriormente determinado. O módulo *SEARCH* a seguir procura na fila a entidade original com o mesmo número da variável *PROCURAR* e retira da fila a entidade original simulando assim um abandono. Depois de liberar o recurso Linhas Telefônicas a entidade sai do sistema.

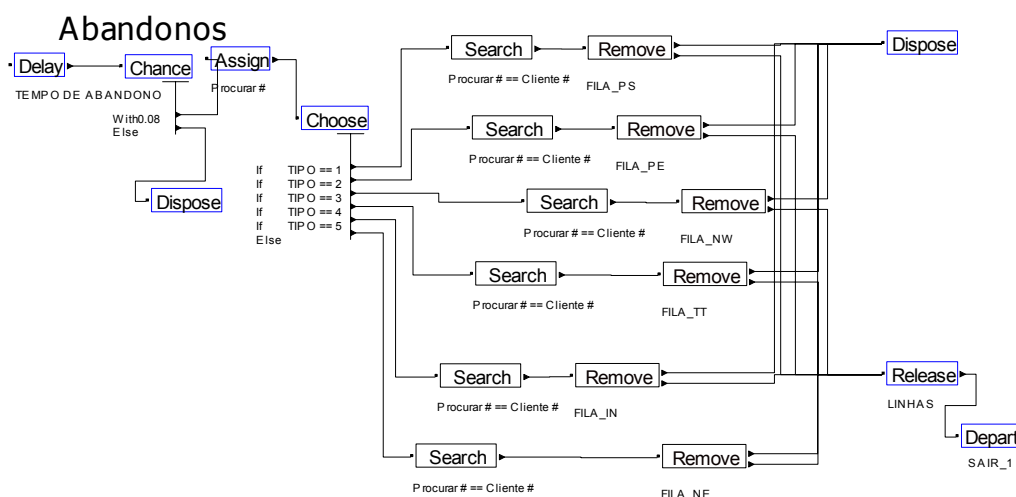


Figura 6: Processo de Abandono de Ligações

A Figura 7 demonstra o modelo completo da central de atendimento.

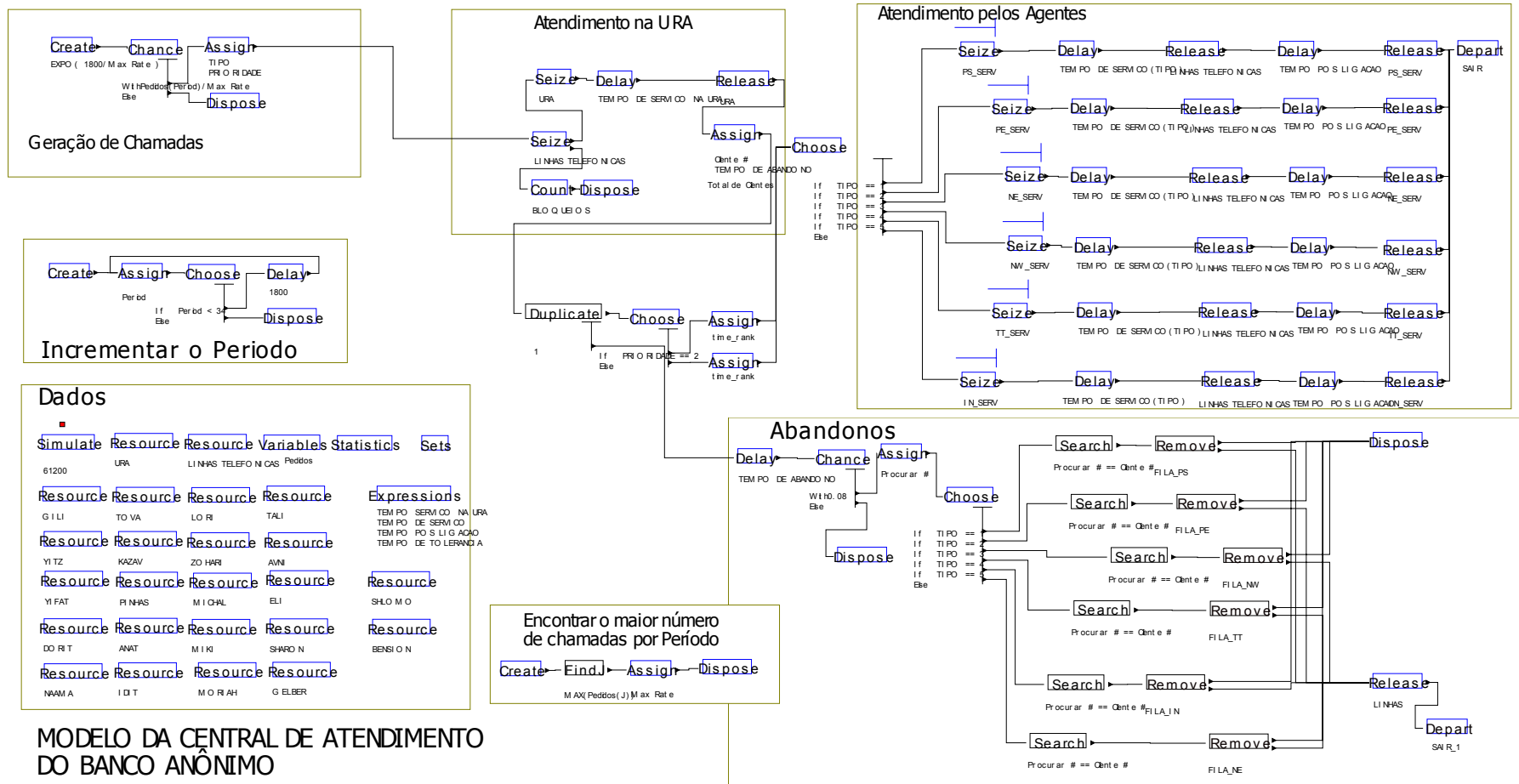


Figura 7: Modelo completo da Central de Atendimento do Banco Anônimo